Improving Accuracy & Efficiency in Document Handling for Business Processes

Introduction to Research in Computer Science
CPSC-59700

Alison Major • 24-August-2021

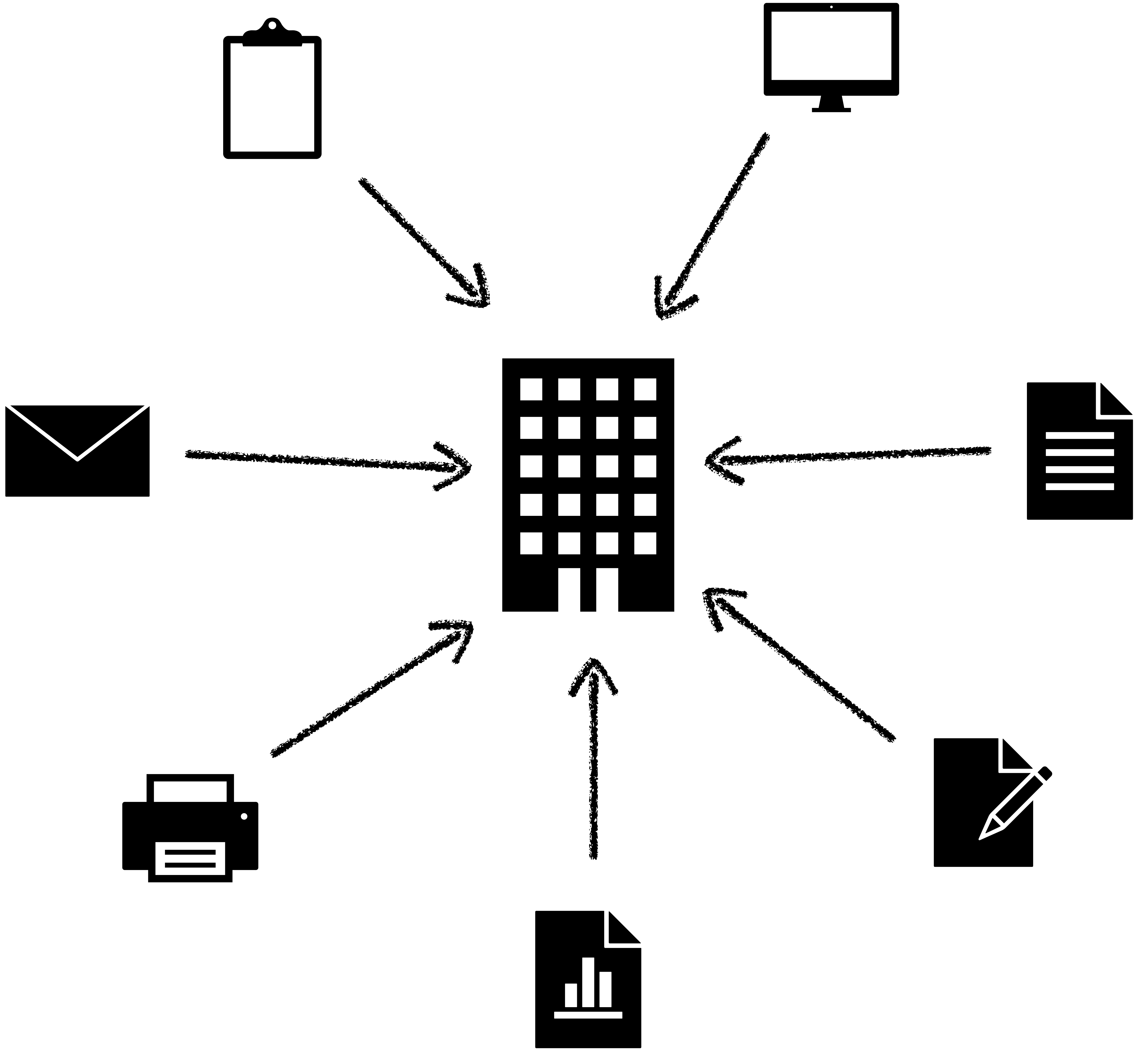
Data comes in many formats .

Manual data extraction is cumbersome and error-prone.

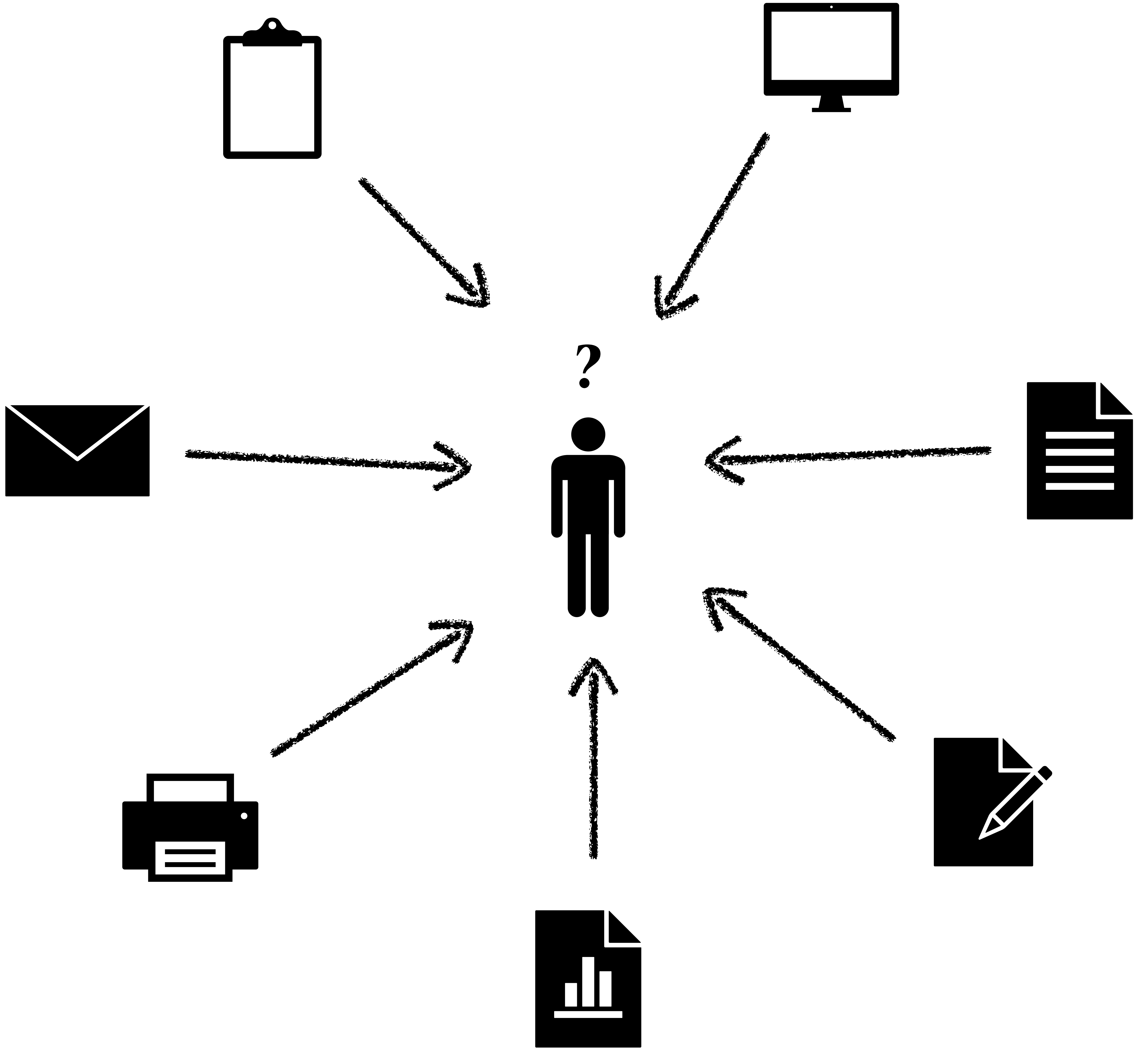
Many tools exist to assist in extracting data,
including OCR, ML, and NLP.

Is there a single method to receive multiple forms of information and automatically pull the desired data?

[Introduction]



[Introduction]



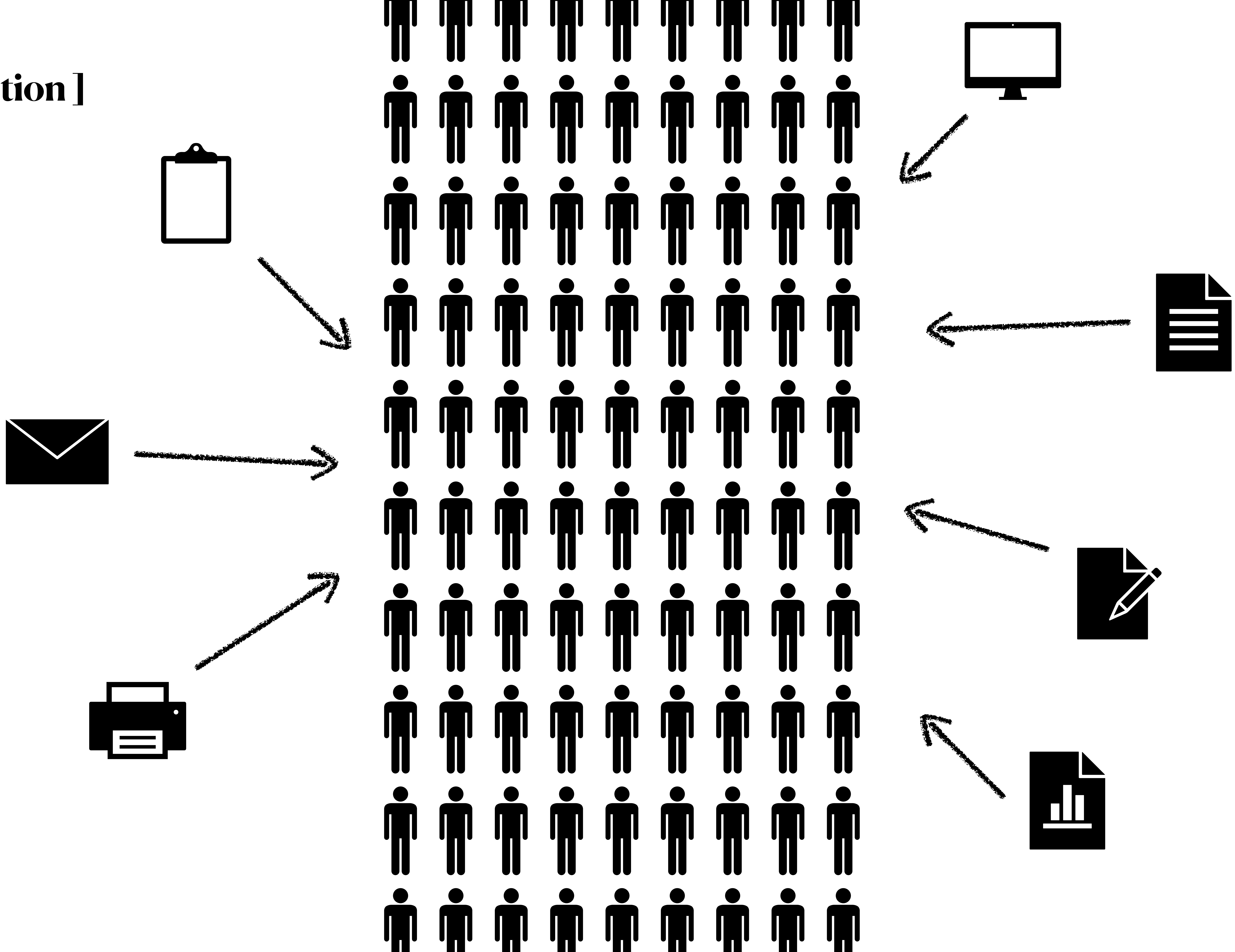
[Introduction]

215 students
30 datasheets
6 data types

Average 10.23 errors

Study from 2009 at UNLV.
Matt Harris. When good info goes bad: The real cost of human data errors, 2014. [Online; accessed 15-August2021].

[Introduction]



[Manual Data Entry]

Direct and indirect cost of manual data entry for global businesses were estimated to be about $2.7 trillion.

2018, Goldman Sachs

James Schneider Ph.D., Bill Schultz, Julia McCrimlisk, Jesse Hulsing, and Ryan M. Nash CFA. B2B: How the next payments frontier will unleash small business. EQUITY RESEARCH, 2018. [Online; accessed 13-August2021].

[Manual Data Entry]

Avoidable Rework for accounting staff of 40: $878,000 per year

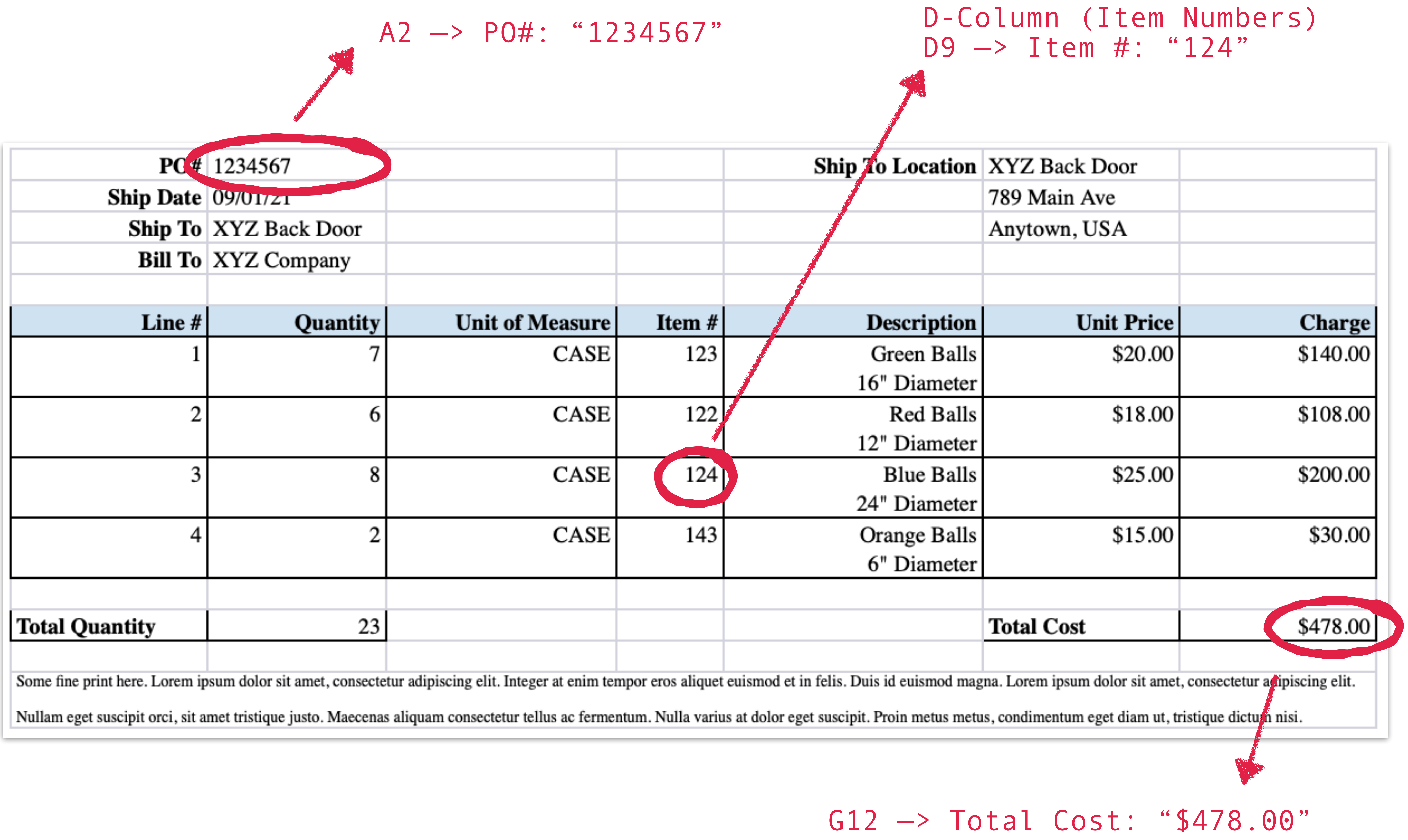
2019, Gartner

Justin Lavelle. Gartner says robotic process automation can save finance departments 25,000 hours of avoidable work annually, 2019. [Online; accessed 13-August2021].

How can we make data extraction simpler?

Rules & Positions

- Machine-readable files can use rules and positions to capture data
- The data is already digital, therefore accuracy is high



A2 —> PO#: “1234567”

D-Column (Item Numbers)
D9 —> Item #: “124”

PO #	1234567			Ship To Location	XYZ Back Door	
Ship Date	09/01/21				789 Main Ave	
Ship To	XYZ Back Door				Anytown, USA	
Bill To	XYZ Company					
Line #	Quantity	Unit of Measure	Item #	Description	Unit Price	Charge
1	7	CASE	123	Green Balls 16" Diameter	$20.00	$140.00
2	6	CASE	122	Red Balls 12" Diameter	$18.00	$108.00
3	8	CASE	124	Blue Balls 24" Diameter	$25.00	$200.00
4	2	CASE	143	Orange Balls 6" Diameter	$15.00	$30.00
Total Quantity	23				Total Cost	$478.00

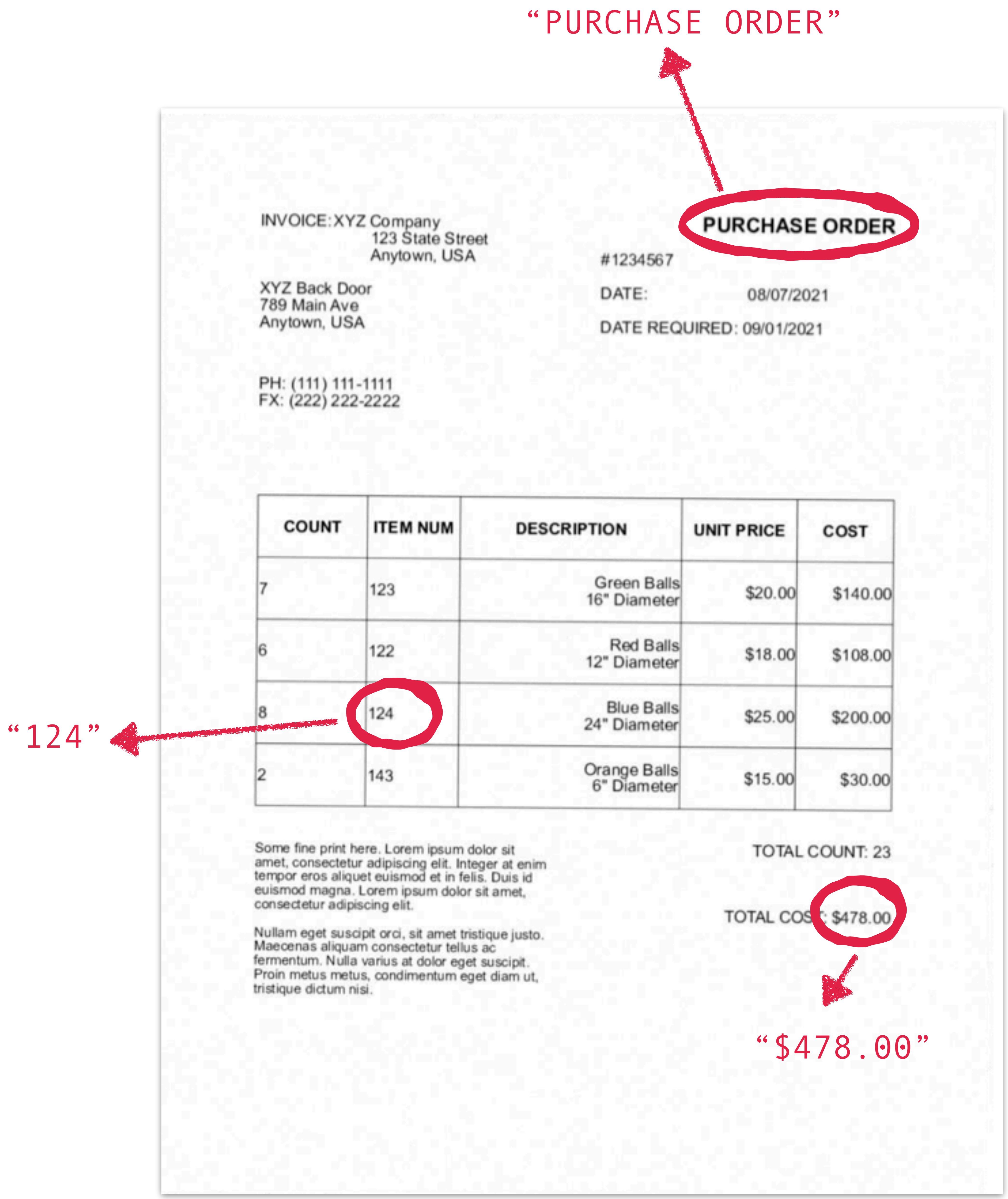
Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

G12 —> Total Cost: “$478.00”

Optical Character Recognition

- Technique that converts images and files into machine-readable data
- Challenges:
  - Orientation & skew
  - Noise or distortion
  - Handwriting or stamps covering printed text



"PURCHASE ORDER"

INVOICE: XYZ Company
123 State Street
Anytown, USA

PURCHASE ORDER

#1234567

XYZ Back Door
789 Main Ave
Anytown, USA

DATE: 08/07/2021

DATE REQUIRED: 09/01/2021

PH: (111) 111-1111
FX: (222) 222-2222

COUNT	ITEM NUM	DESCRIPTION	UNIT PRICE	COST
7	123	Green Balls 16" Diameter	$20.00	$140.00
6	122	Red Balls 12" Diameter	$18.00	$108.00
8	124	Blue Balls 24" Diameter	$25.00	$200.00
2	143	Orange Balls 6" Diameter	$15.00	$30.00

"124"

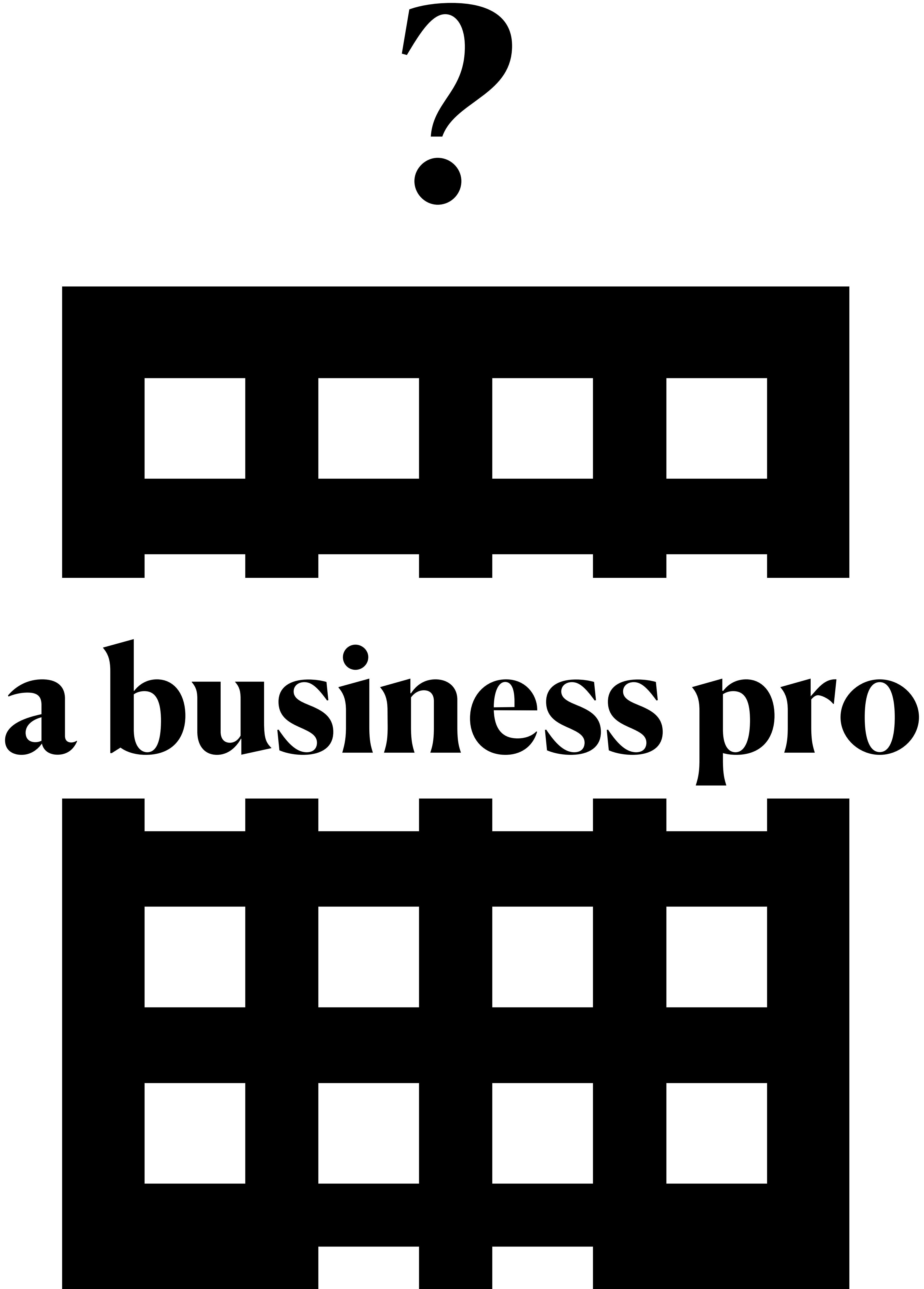
Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

TOTAL COUNT: 23

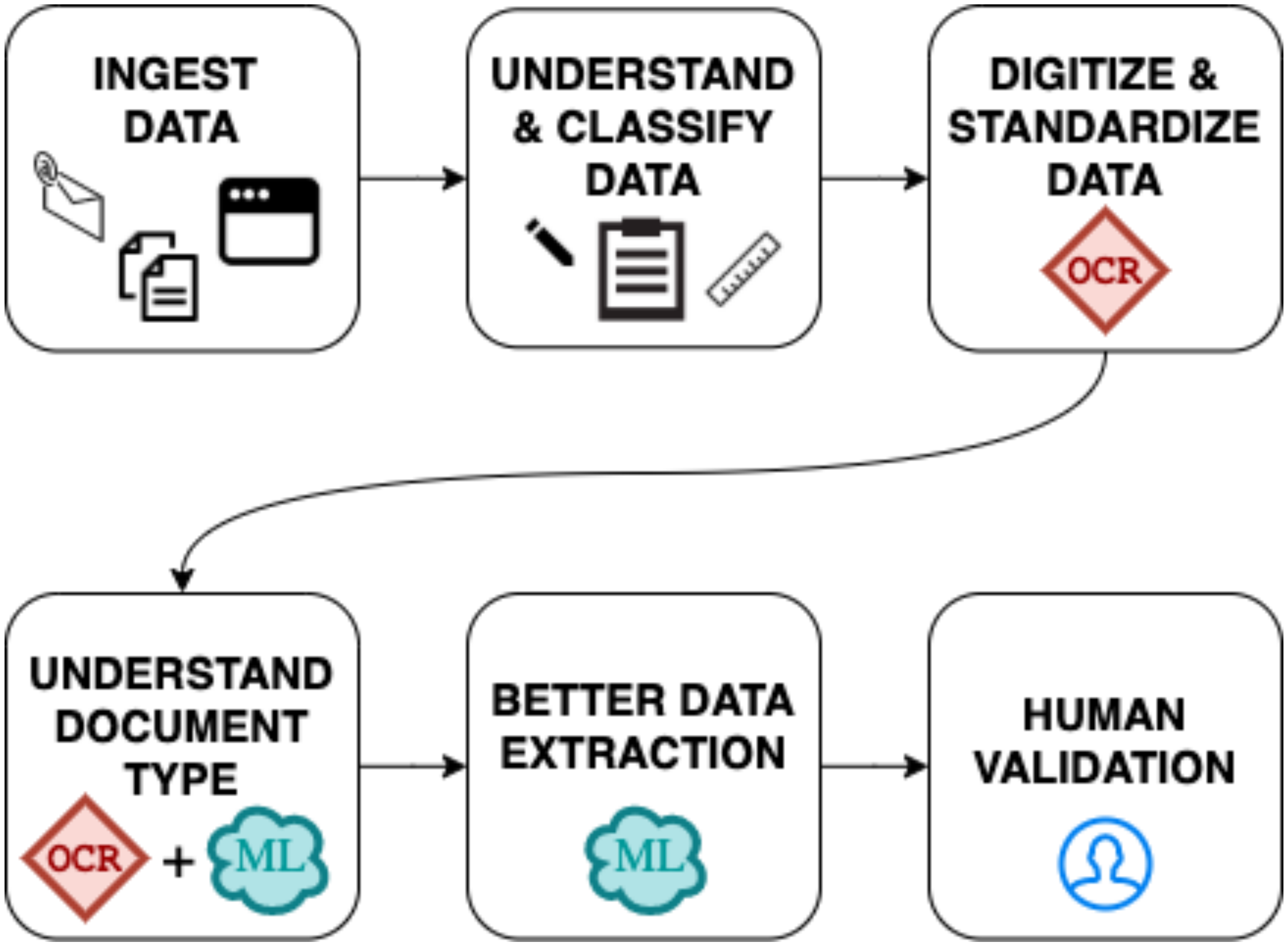
TOTAL COST: $478.00

"$478.00"



How does a business process data?

[Processing Data]



INGEST DATA

UNDERSTAND & CLASSIFY DATA

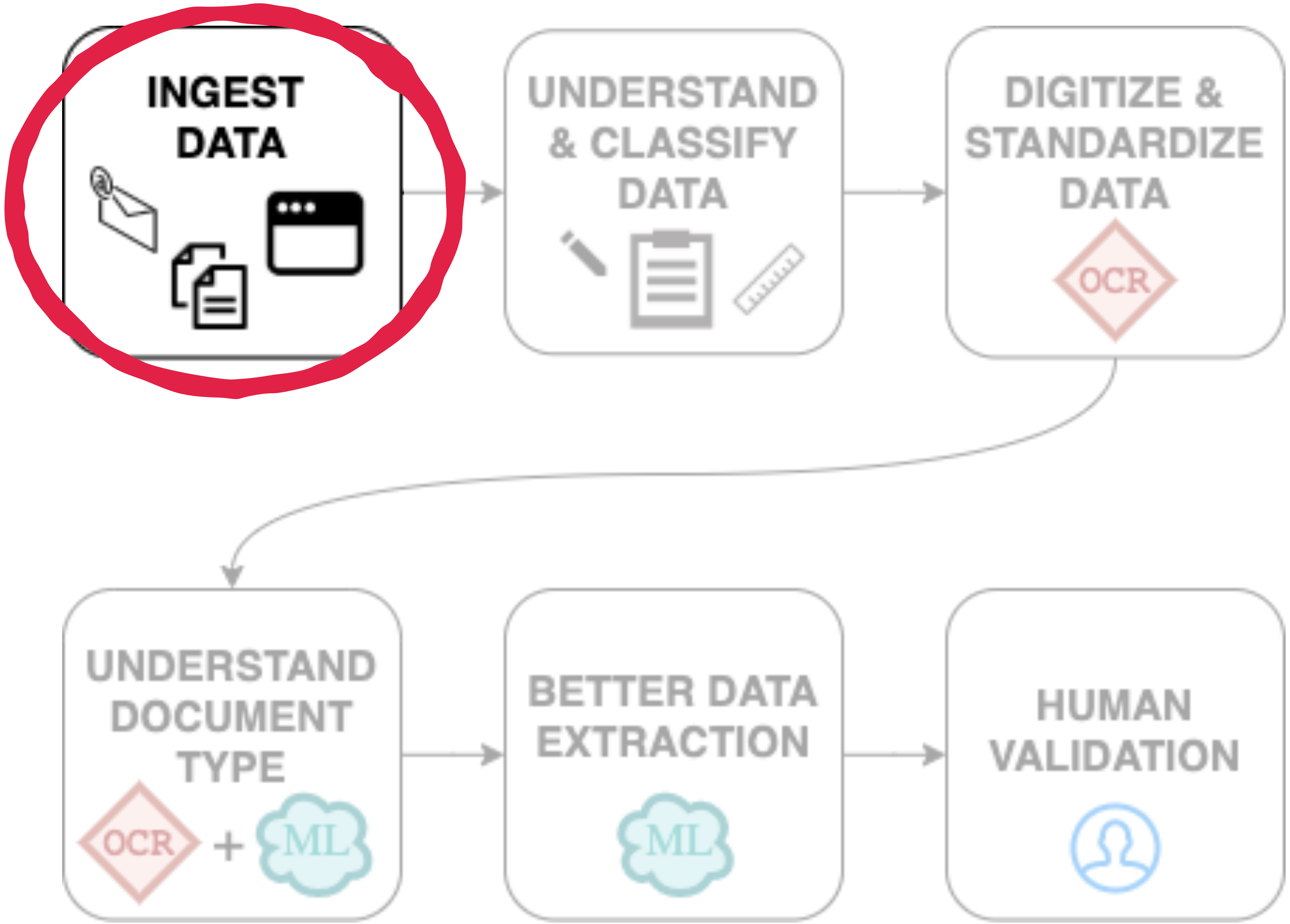
DIGITIZE & STANDARDIZE DATA

UNDERSTAND DOCUMENT TYPE

BETTER DATA EXTRACTION

HUMAN VALIDATION

[Processing Data]



INGEST DATA

UNDERSTAND & CLASSIFY DATA

DIGITIZE & STANDARDIZE DATA

OCR

UNDERSTAND DOCUMENT TYPE

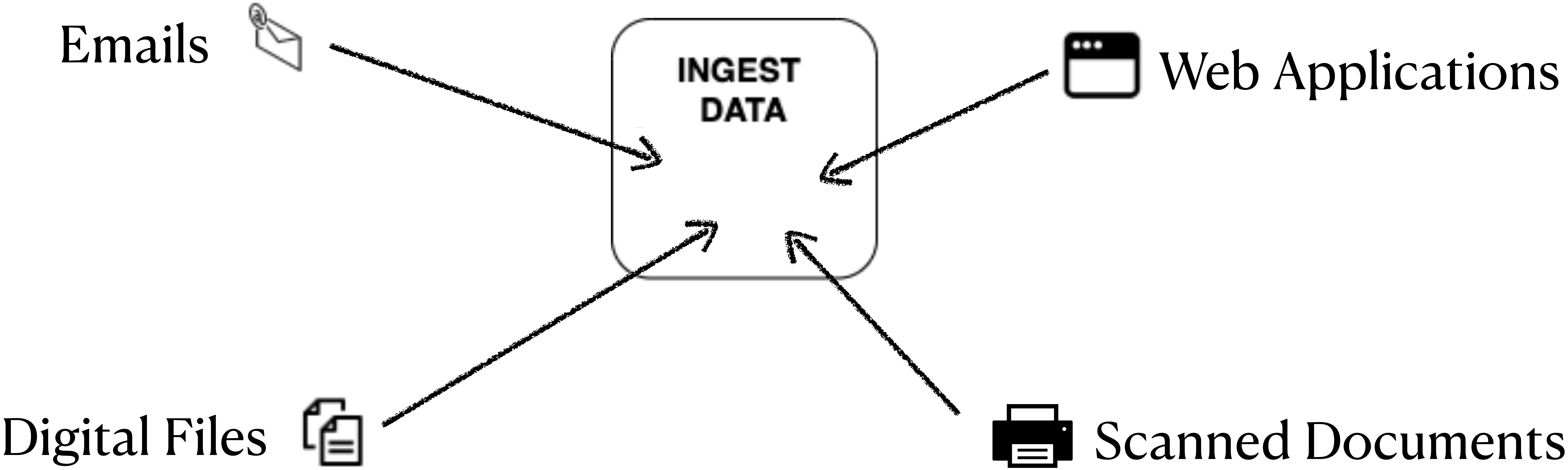
OCR + ML

BETTER DATA EXTRACTION

ML

HUMAN VALIDATION

[Processing Data]



INGEST DATA

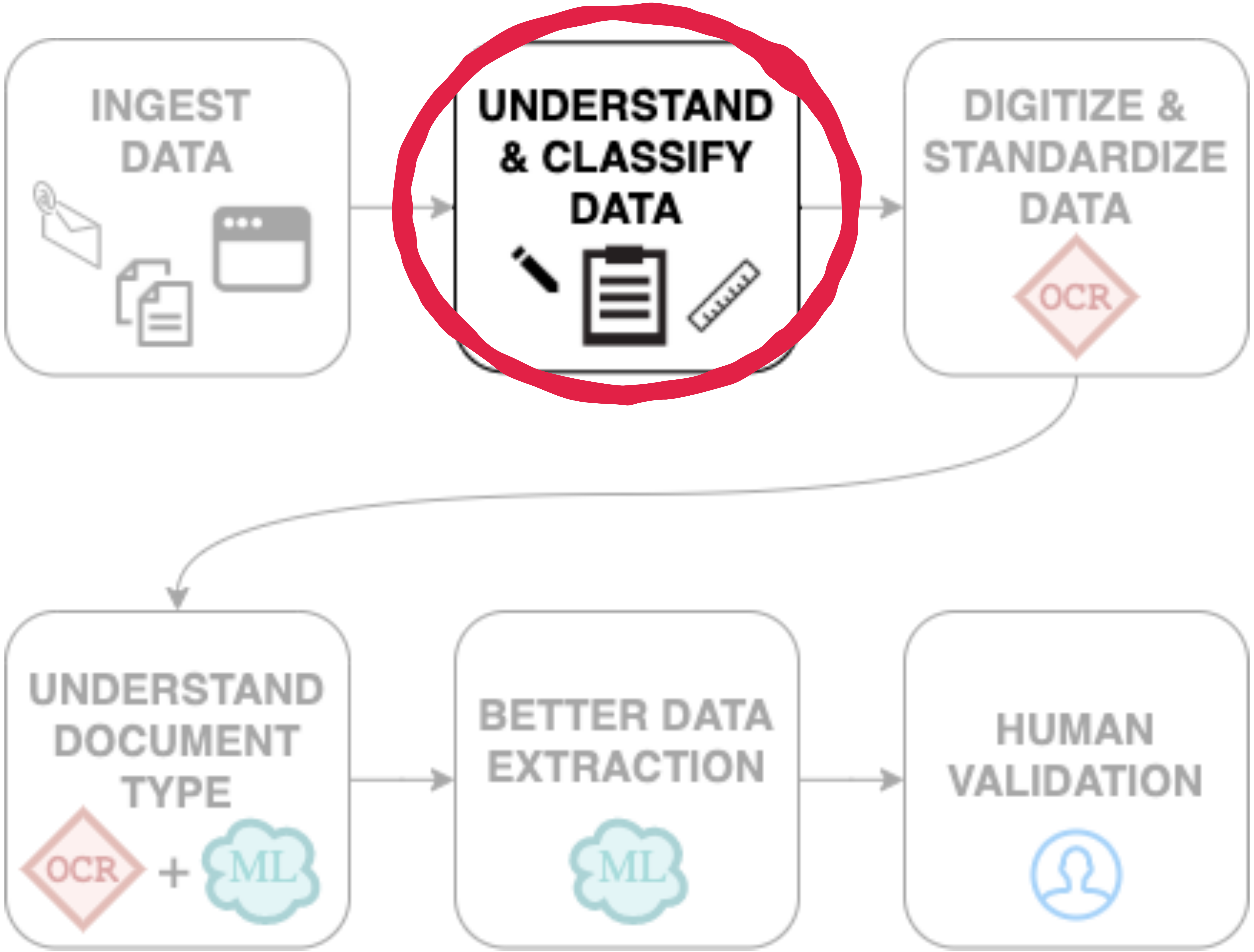
Emails

Web Applications

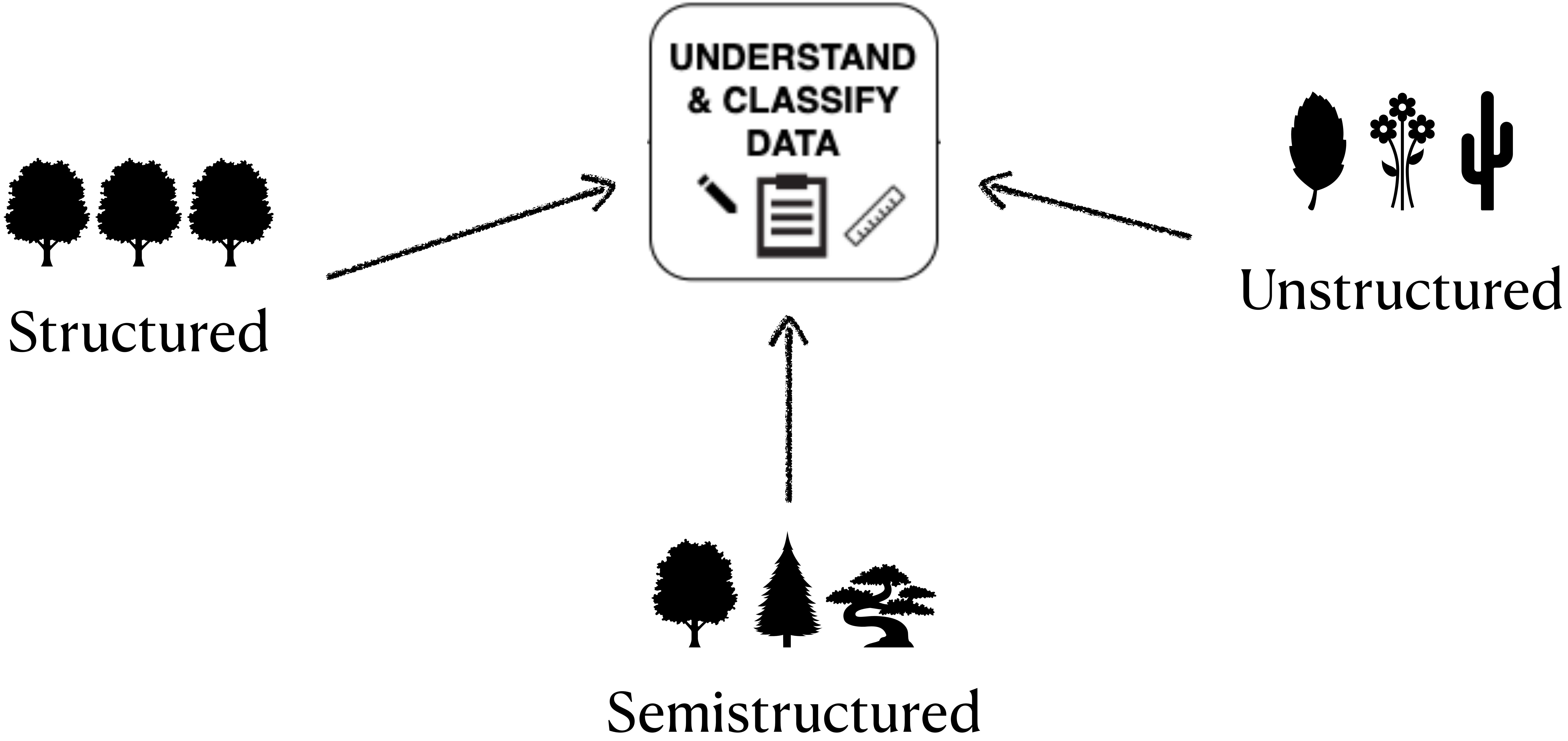
Digital Files

Scanned Documents

[Processing Data]



[Processing Data]



UNDERSTAND & CLASSIFY DATA

Structured

Unstructured

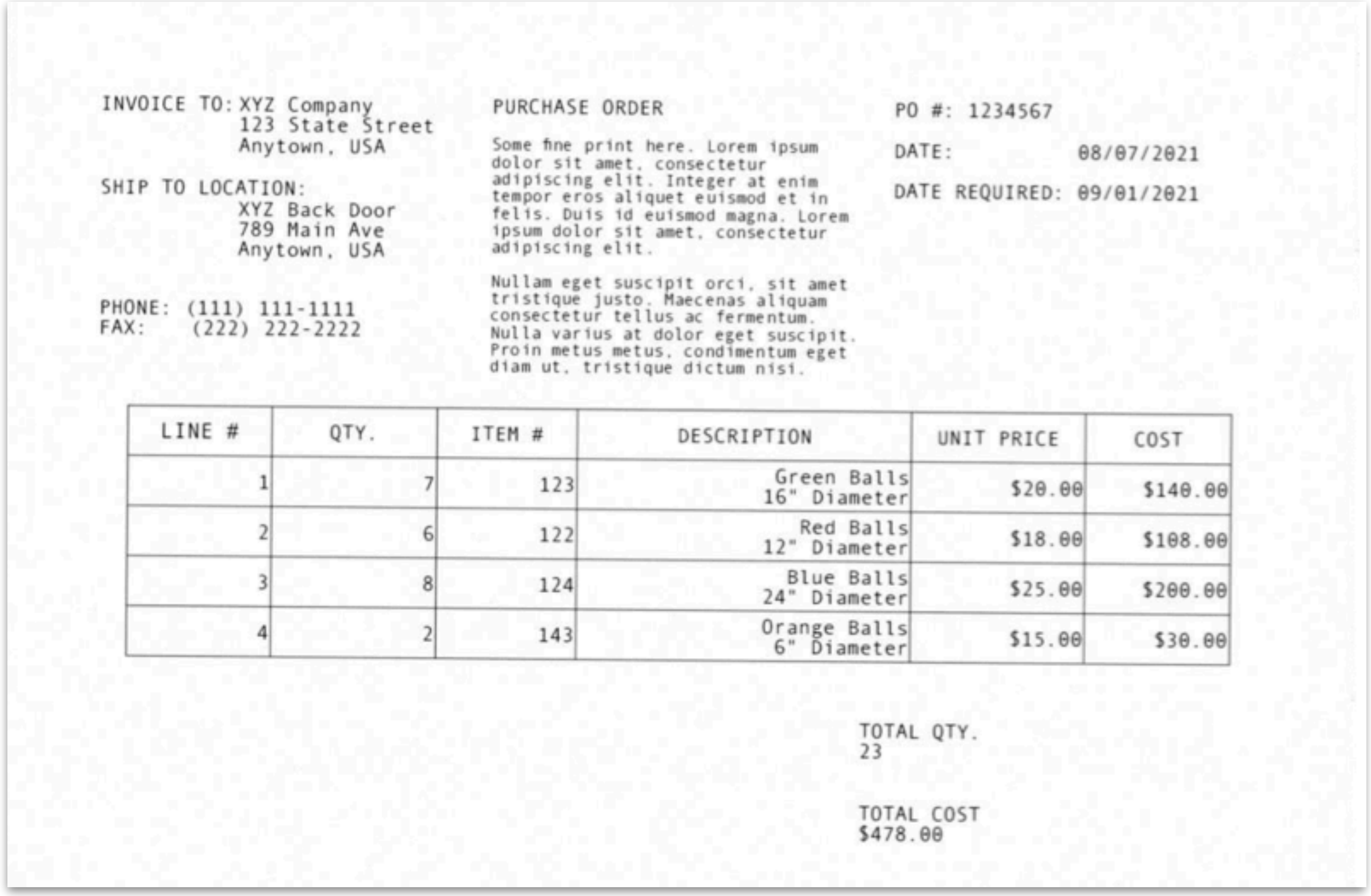
Semistructured

[Processing Data]

Structured

- Formatted in a consistent template
- Tables with values in predictable positions
- Same layout for every document

Layout 1



INVOICE TO: XYZ Company
123 State Street
Anytown, USA

SHIP TO LOCATION:
XYZ Back Door
789 Main Ave
Anytown, USA

PHONE: (111) 111-1111
FAX: (222) 222-2222

PURCHASE ORDER

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

PO #: 1234567

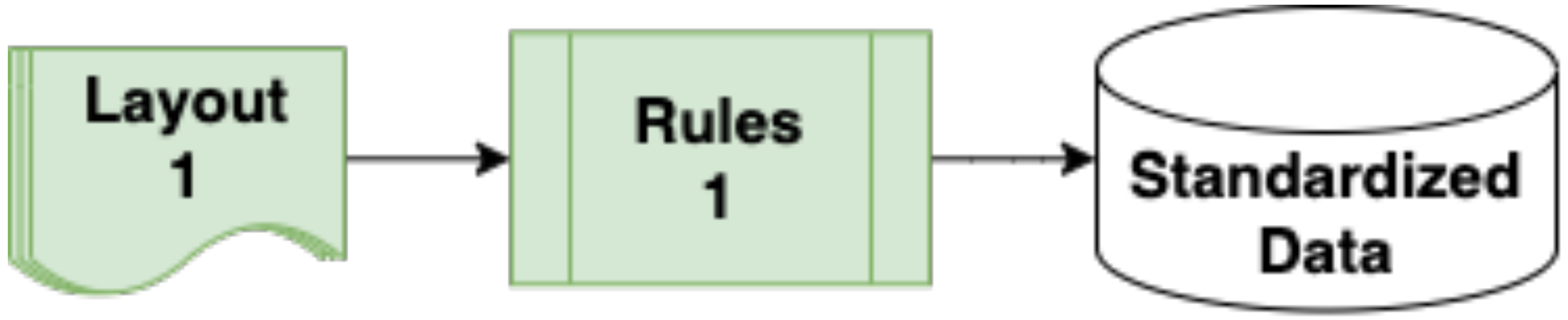
DATE: 08/07/2021

DATE REQUIRED: 09/01/2021

LINE #	QTY.	ITEM #	DESCRIPTION	UNIT PRICE	COST
1	7	123	Green Balls 16" Diameter	$20.00	$140.00
2	6	122	Red Balls 12" Diameter	$18.00	$108.00
3	8	124	Blue Balls 24" Diameter	$25.00	$200.00
4	2	143	Orange Balls 6" Diameter	$15.00	$30.00

TOTAL QTY.
23

TOTAL COST
$478.00

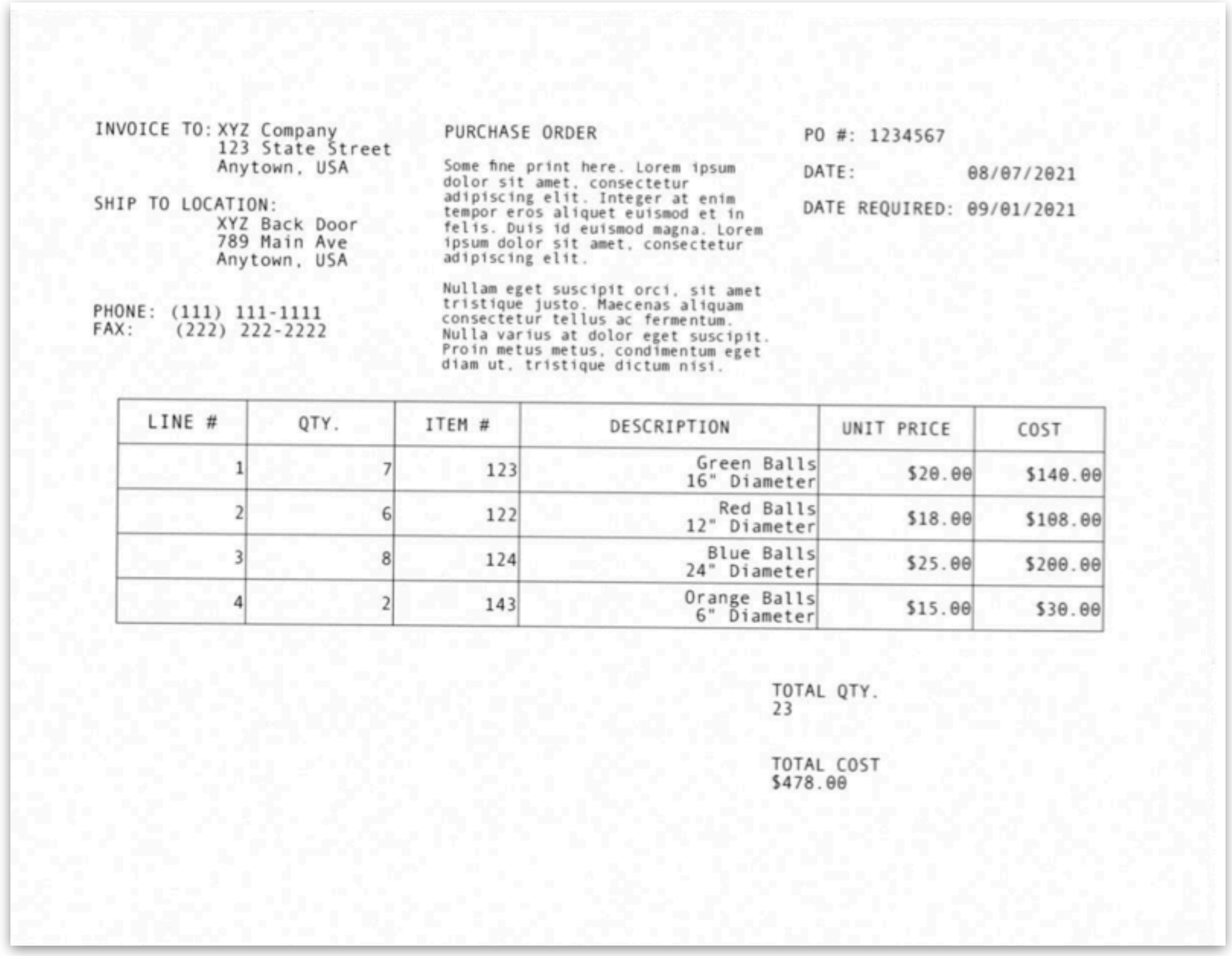


[Processing Data]

Semistructured

- Same key-value pairs
- Different layouts for the information
- Positions of values vary
- Using the same rules for a different layout results in low accuracy for the "wrong" layout

Layout 1



INVOICE TO: XYZ Company
123 State Street
Anytown, USA

SHIP TO LOCATION:
XYZ Back Door
789 Main Ave
Anytown, USA

PHONE: (111) 111-1111
FAX: (222) 222-2222

PURCHASE ORDER

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

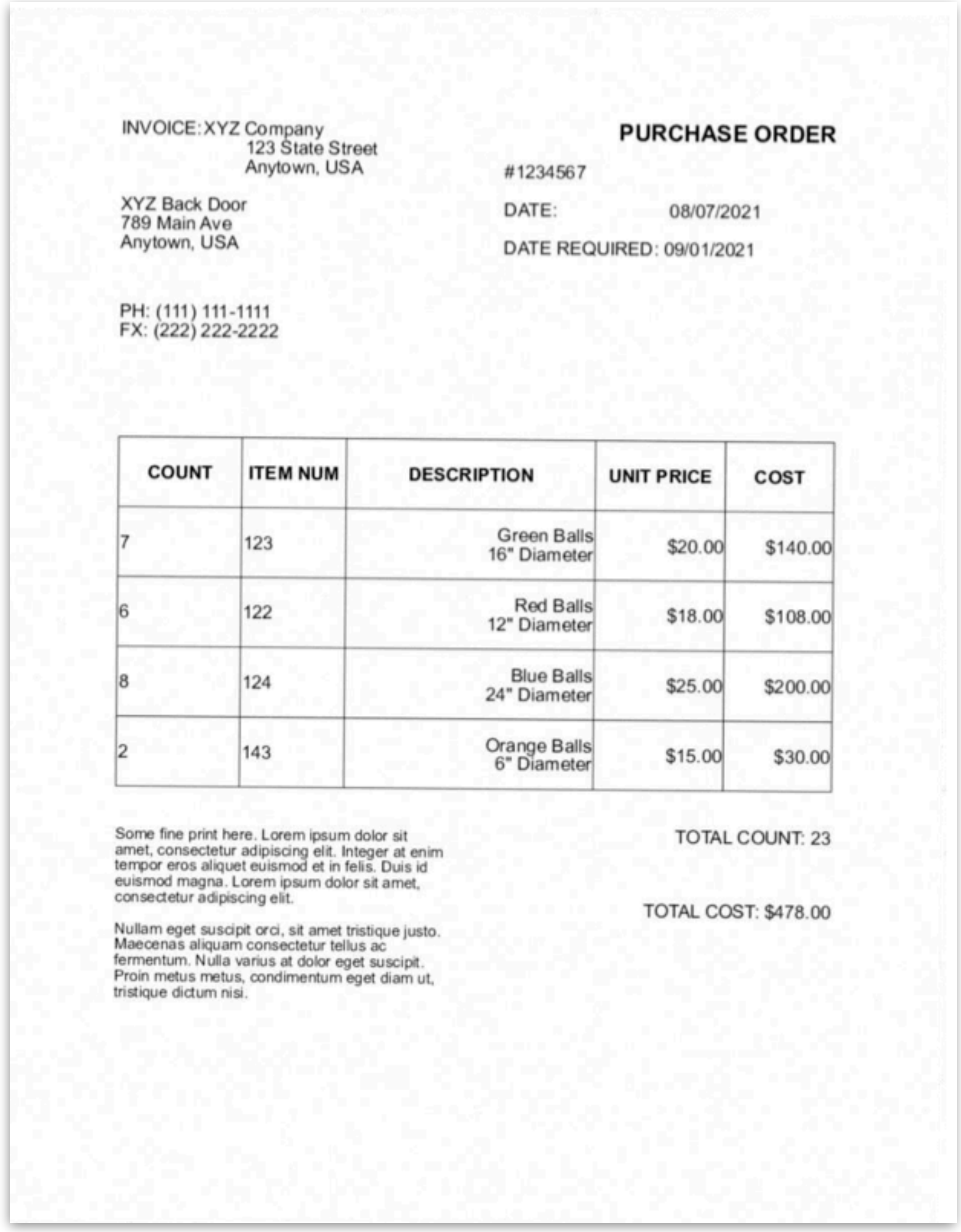
PO #: 1234567
DATE: 08/07/2021
DATE REQUIRED: 09/01/2021

LINE #	QTY.	ITEM #	DESCRIPTION	UNIT PRICE	COST
1	7	123	Green Balls 16" Diameter	$20.00	$140.00
2	6	122	Red Balls 12" Diameter	$18.00	$108.00
3	8	124	Blue Balls 24" Diameter	$25.00	$200.00
4	2	143	Orange Balls 6" Diameter	$15.00	$30.00

TOTAL QTY.
23

TOTAL COST
$478.00

Layout 2



INVOICE: XYZ Company
123 State Street
Anytown, USA

XYZ Back Door
789 Main Ave
Anytown, USA

PH: (111) 111-1111
FX: (222) 222-2222

PURCHASE ORDER

#1234567
DATE: 08/07/2021
DATE REQUIRED: 09/01/2021

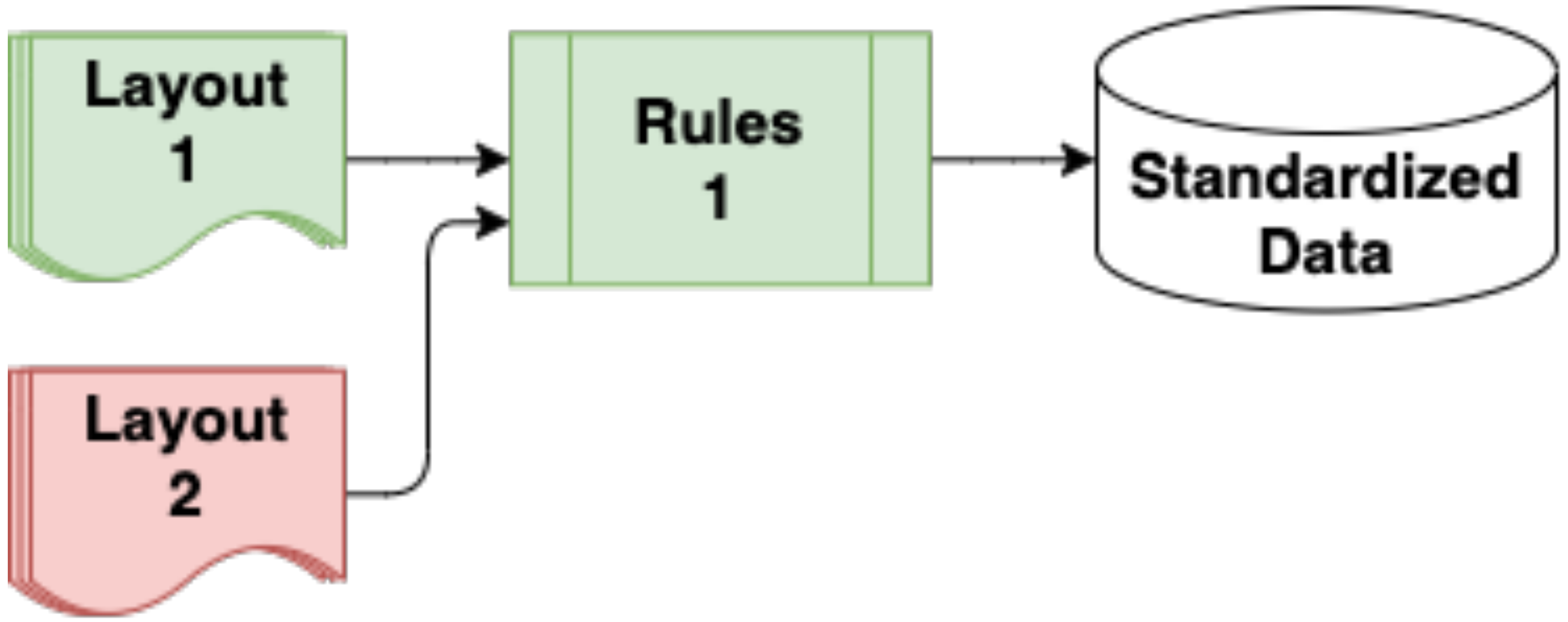
COUNT	ITEM NUM	DESCRIPTION	UNIT PRICE	COST
7	123	Green Balls 16" Diameter	$20.00	$140.00
6	122	Red Balls 12" Diameter	$18.00	$108.00
8	124	Blue Balls 24" Diameter	$25.00	$200.00
2	143	Orange Balls 6" Diameter	$15.00	$30.00

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

TOTAL COUNT: 23

TOTAL COST: $478.00



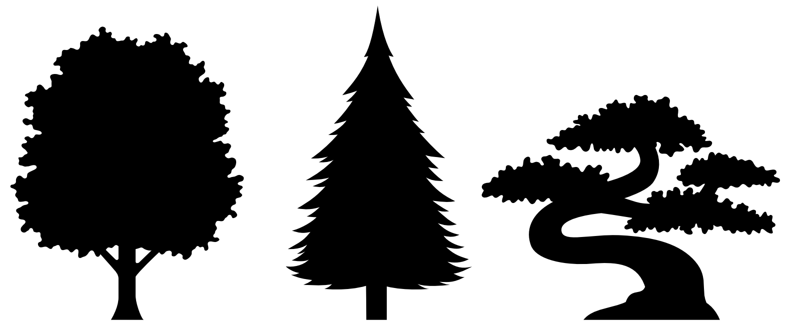
Layout 1

Layout 2

Rules 1

Standardized Data

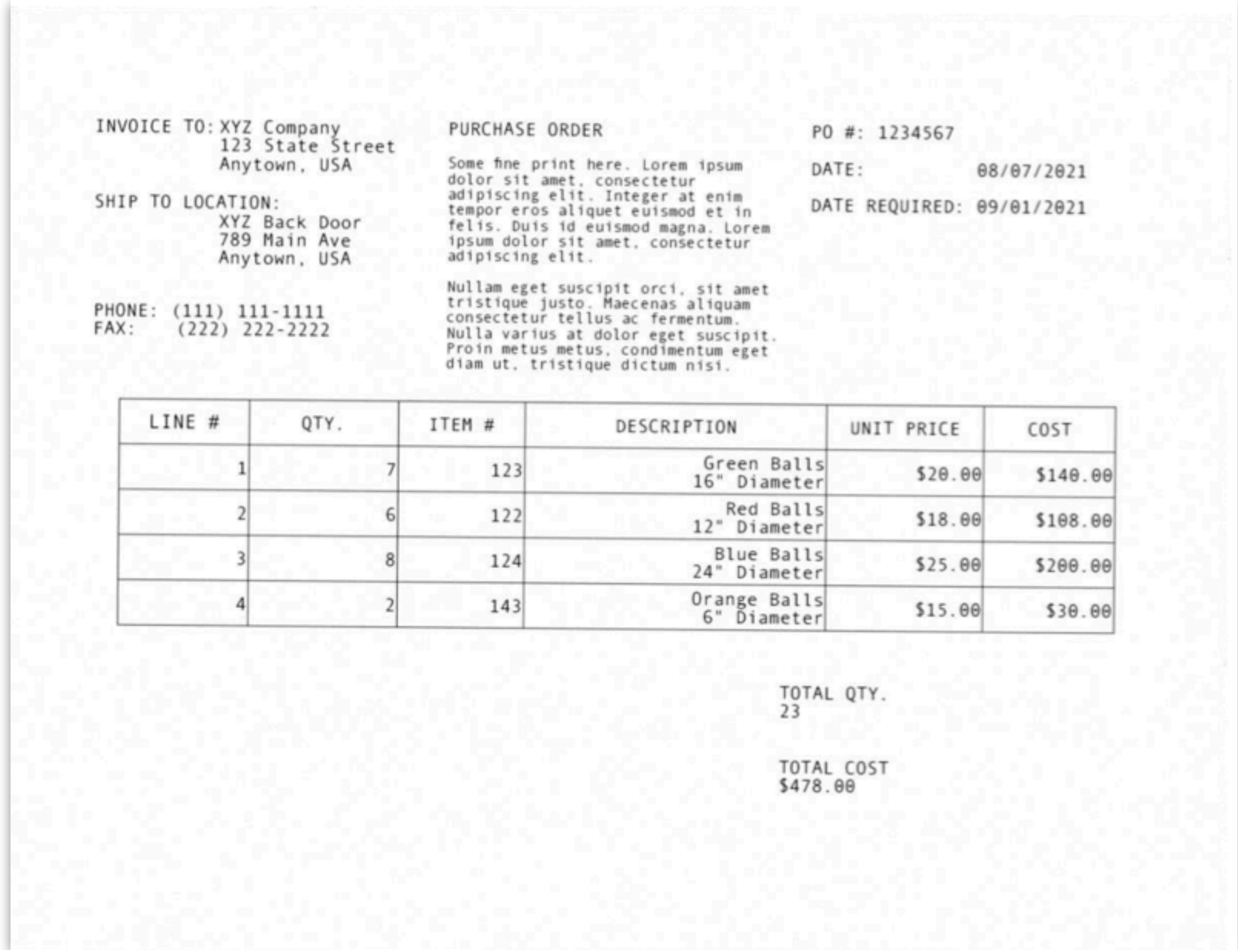
[Processing Data]



Semistructured

- Create rules for each layout
- Scaling...
  requires template
  and rules management!

Layout 1



INVOICE TO: XYZ Company
123 State Street
Anytown, USA

SHIP TO LOCATION:
XYZ Back Door
789 Main Ave
Anytown, USA

PHONE: (111) 111-1111
FAX: (222) 222-2222

PURCHASE ORDER

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

PO #: 1234567

DATE: 08/07/2021

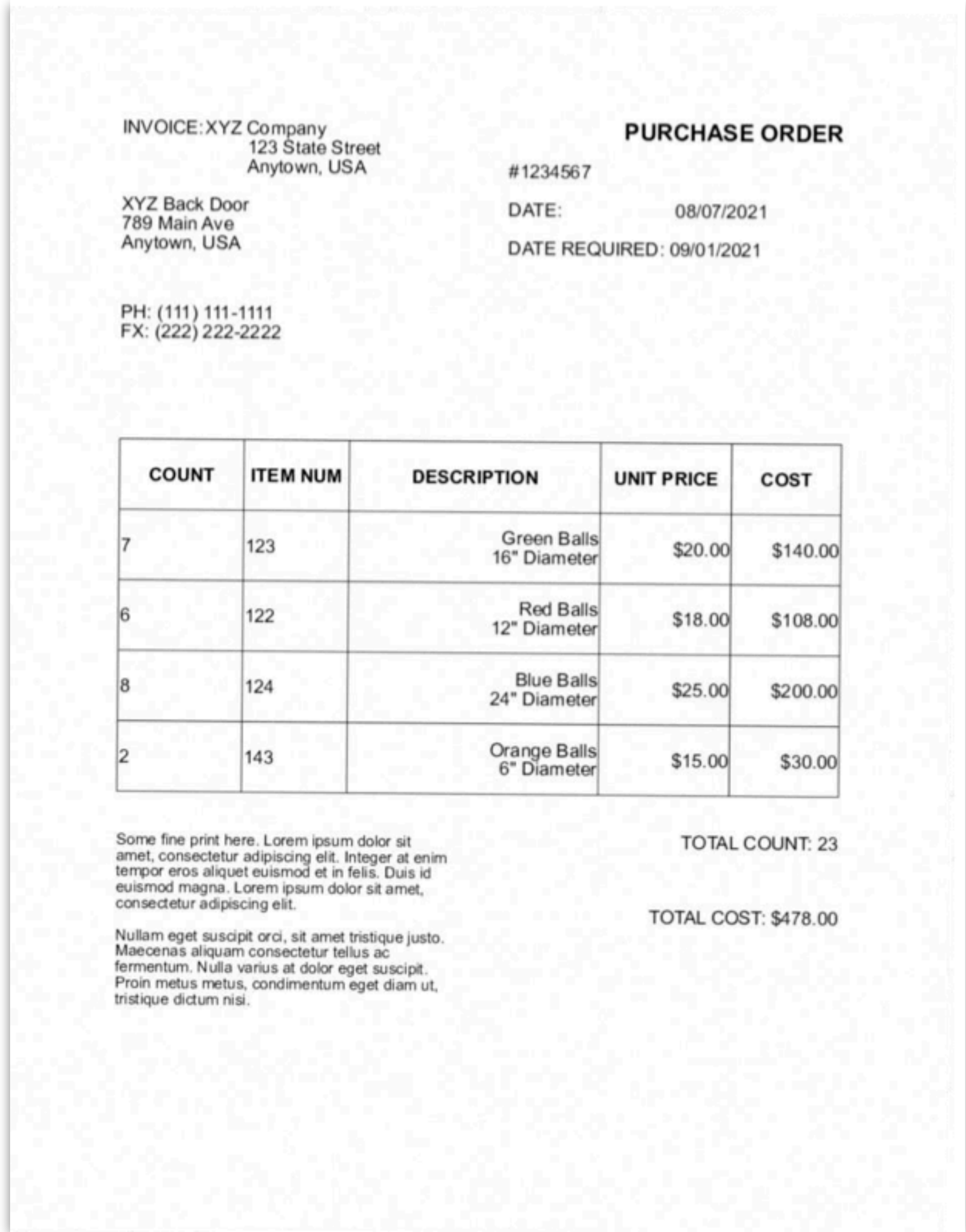
DATE REQUIRED: 09/01/2021

LINE #	QTY.	ITEM #	DESCRIPTION	UNIT PRICE	COST
1	7	123	Green Balls 16" Diameter	$20.00	$140.00
2	6	122	Red Balls 12" Diameter	$18.00	$108.00
3	8	124	Blue Balls 24" Diameter	$25.00	$200.00
4	2	143	Orange Balls 6" Diameter	$15.00	$30.00

TOTAL QTY.
23

TOTAL COST
$478.00

Layout 2



INVOICE: XYZ Company
123 State Street
Anytown, USA

XYZ Back Door
789 Main Ave
Anytown, USA

PH: (111) 111-1111
FX: (222) 222-2222

PURCHASE ORDER

#1234567

DATE: 08/07/2021

DATE REQUIRED: 09/01/2021

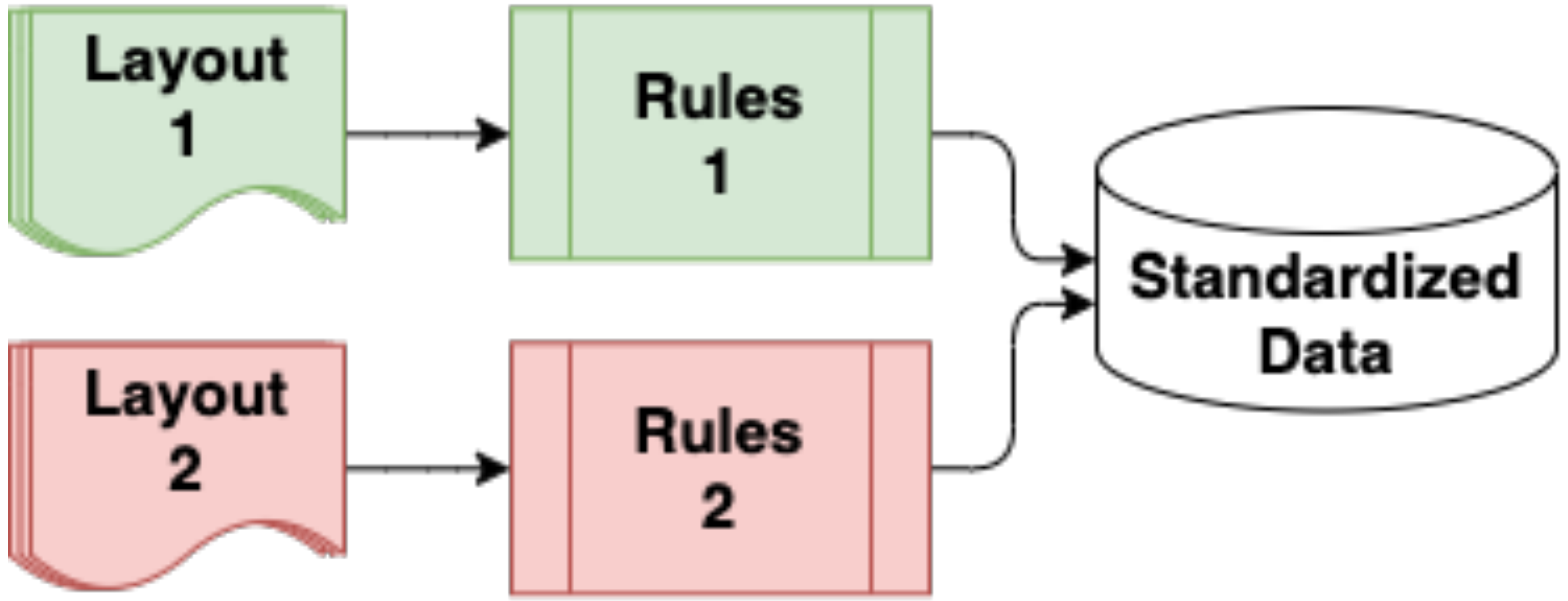
COUNT	ITEM NUM	DESCRIPTION	UNIT PRICE	COST
7	123	Green Balls 16" Diameter	$20.00	$140.00
6	122	Red Balls 12" Diameter	$18.00	$108.00
8	124	Blue Balls 24" Diameter	$25.00	$200.00
2	143	Orange Balls 6" Diameter	$15.00	$30.00

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

TOTAL COUNT: 23

TOTAL COST: $478.00



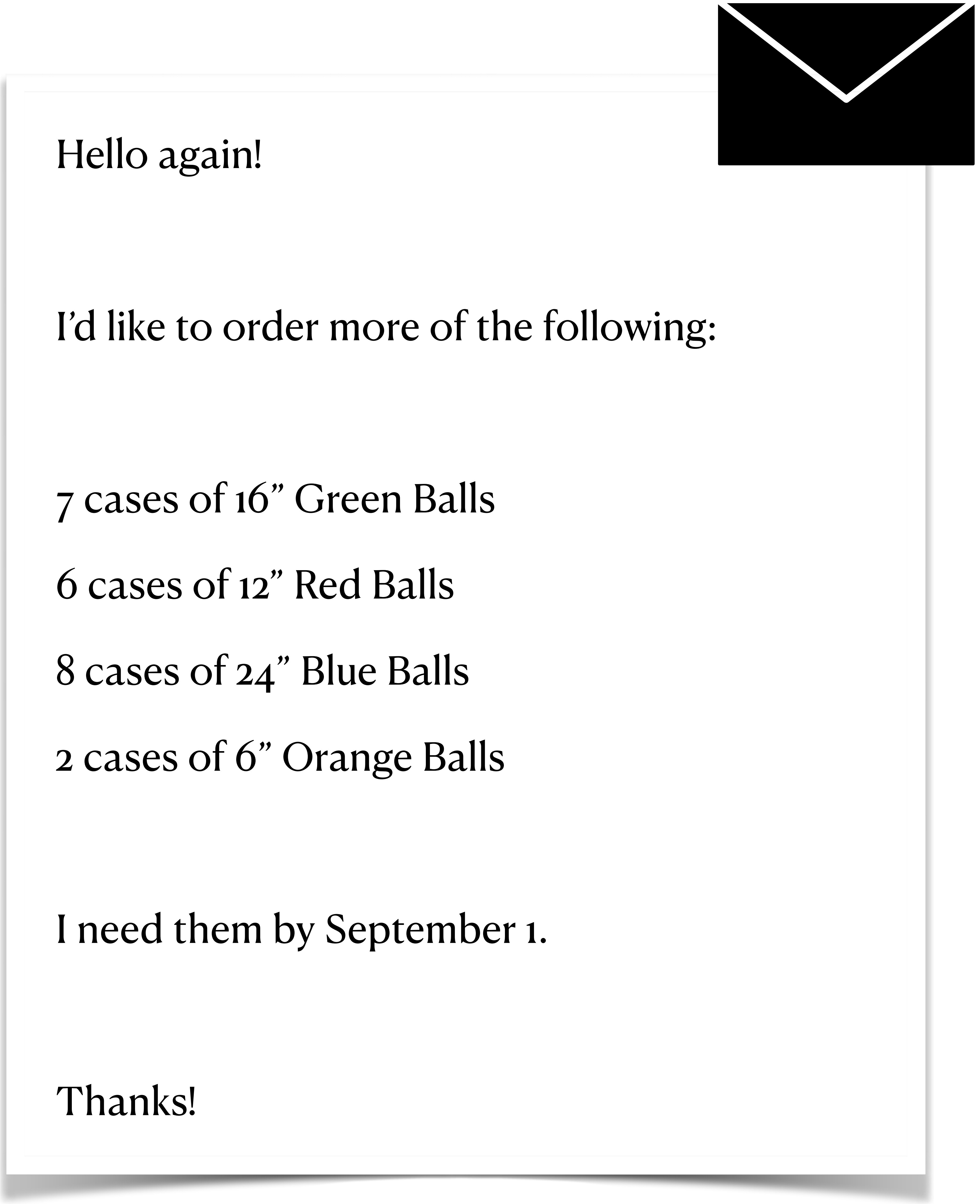
Layout 1 → Rules 1 → Standardized Data

Layout 2 → Rules 2 → Standardized Data

[Processing Data]

Unstructured

- No key-value pairs
- Free-flowing
- Lacks consistency
- May contain handwriting
- Consider NLP for extraction



Hello again!

I'd like to order more of the following:

7 cases of 16" Green Balls

6 cases of 12" Red Balls

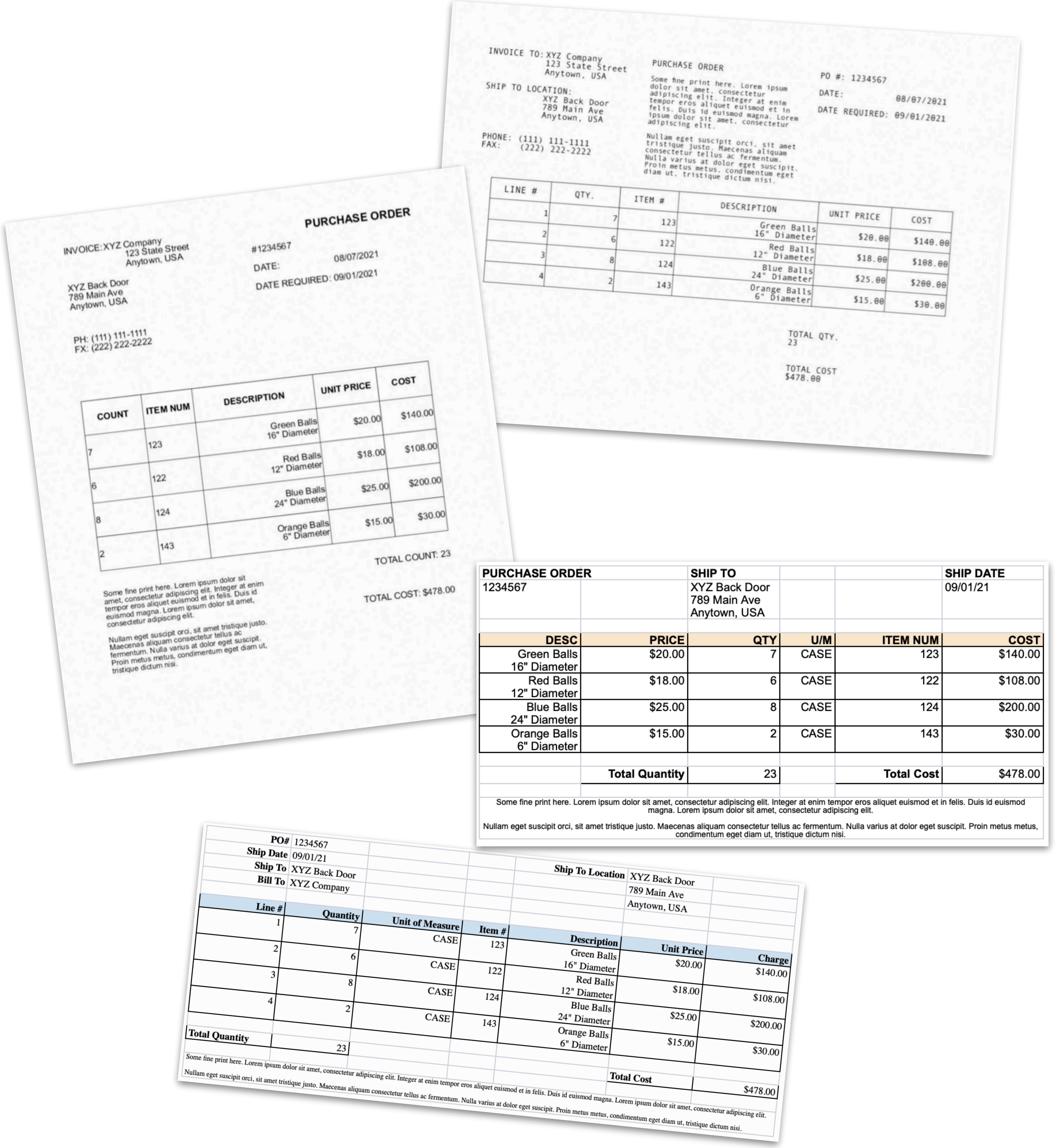
8 cases of 24" Blue Balls

2 cases of 6" Orange Balls

I need them by September 1.

Thanks!

[Processing Data]



INVOICE TO: XYZ Company
123 State Street
Anytown, USA

SHIP TO LOCATION:
XYZ Back Door
789 Main Ave
Anytown, USA

PHONE: (111) 111-1111
FAX: (222) 222-2222

PURCHASE ORDER

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

PO #: 1234567
DATE: 08/07/2021
DATE REQUIRED: 09/01/2021

LINE #	QTY.	ITEM #	DESCRIPTION	UNIT PRICE	COST
1	7	123	Green Balls 16" Diameter	$20.00	$140.00
2	6	122	Red Balls 12" Diameter	$18.00	$108.00
3	8	124	Blue Balls 24" Diameter	$25.00	$200.00
4	2	143	Orange Balls 6" Diameter	$15.00	$30.00

TOTAL QTY.
23

TOTAL COST
$478.00

INVOICE: XYZ Company
123 State Street
Anytown, USA

XYZ Back Door
789 Main Ave
Anytown, USA

PH: (111) 111-1111
FX: (222) 222-2222

PURCHASE ORDER

#1234567
DATE: 08/07/2021
DATE REQUIRED: 09/01/2021

COUNT	ITEM NUM	DESCRIPTION	UNIT PRICE	COST
7	123	Green Balls 16" Diameter	$20.00	$140.00
6	122	Red Balls 12" Diameter	$18.00	$108.00
8	124	Blue Balls 24" Diameter	$25.00	$200.00
2	143	Orange Balls 6" Diameter	$15.00	$30.00

TOTAL COUNT: 23

TOTAL COST: $478.00

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

PURCHASE ORDER
1234567

SHIP TO
XYZ Back Door
789 Main Ave
Anytown, USA

SHIP DATE
09/01/21

DESC	PRICE	QTY	U/M	ITEM NUM	COST
Green Balls 16" Diameter	$20.00	7	CASE	123	$140.00
Red Balls 12" Diameter	$18.00	6	CASE	122	$108.00
Blue Balls 24" Diameter	$25.00	8	CASE	124	$200.00
Orange Balls 6" Diameter	$15.00	2	CASE	143	$30.00
	Total Quantity	23		Total Cost	$478.00

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

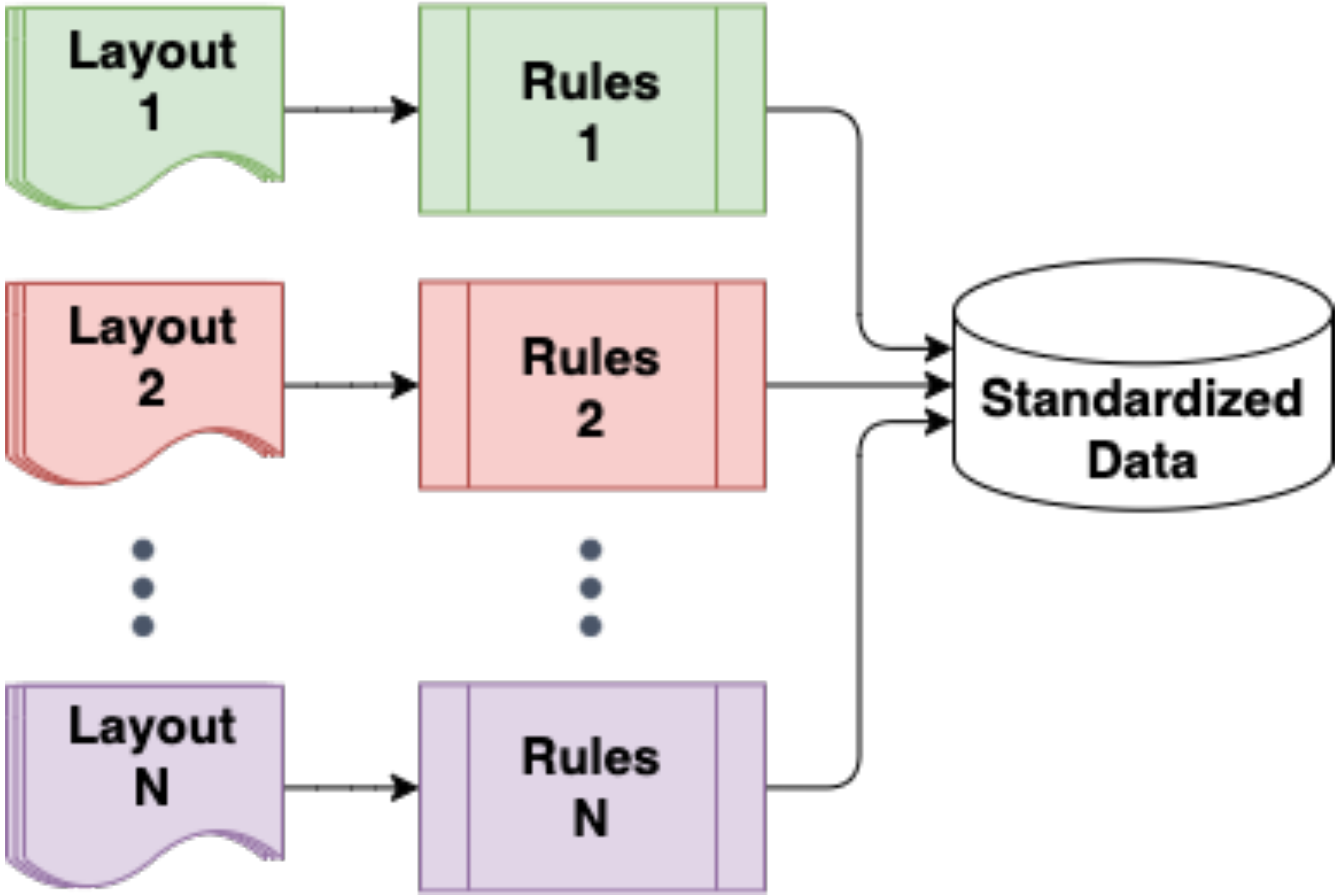
PO# 1234567
Ship Date 09/01/21
Ship To XYZ Back Door
Bill To XYZ Company

Ship To Location XYZ Back Door
789 Main Ave
Anytown, USA

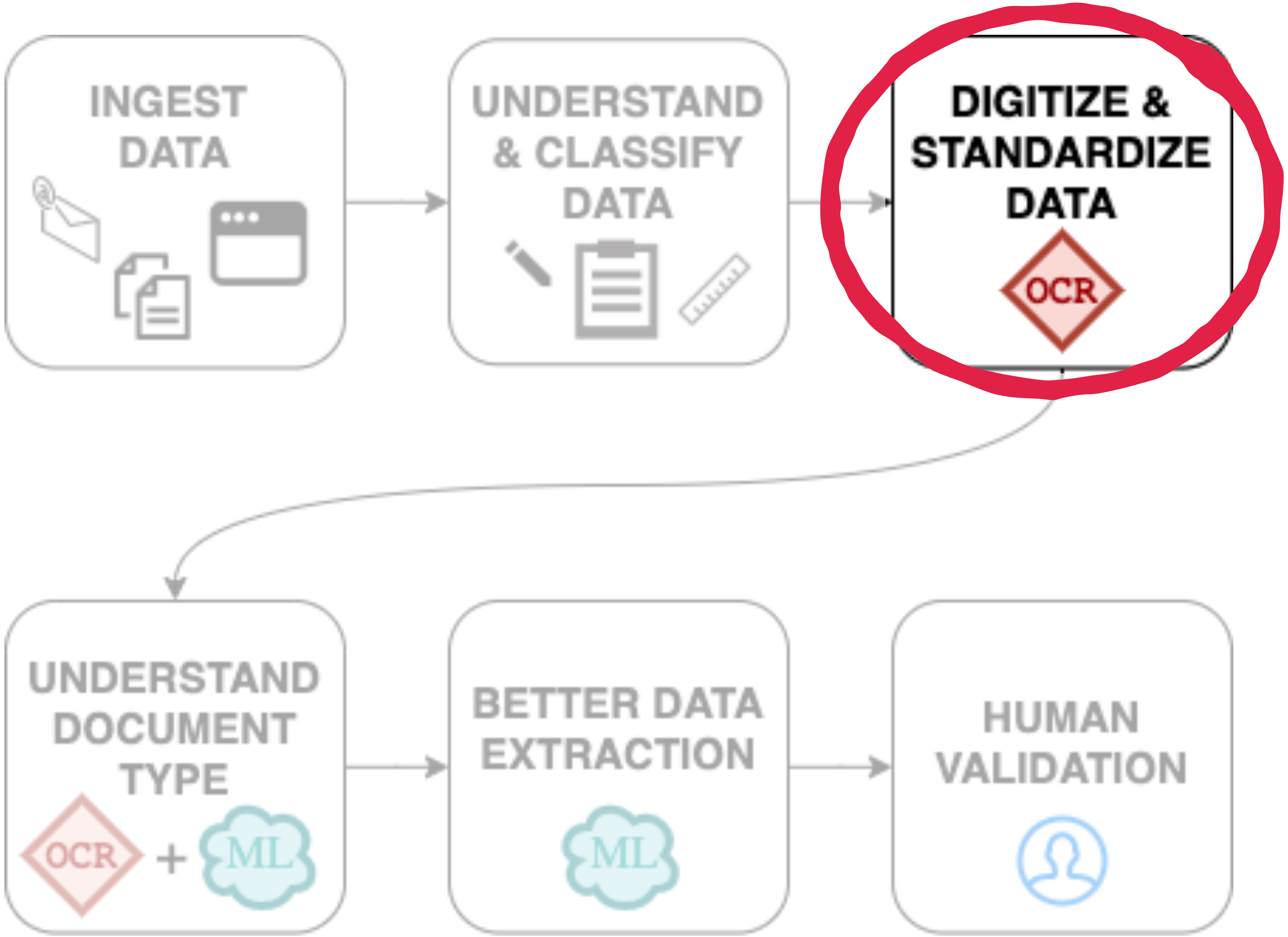
Line #	Quantity	Unit of Measure	Item #	Description	Unit Price	Charge
1	7	CASE	123	Green Balls 16" Diameter	$20.00	$140.00
2	6	CASE	122	Red Balls 12" Diameter	$18.00	$108.00
3	8	CASE	124	Blue Balls 24" Diameter	$25.00	$200.00
4	2	CASE	143	Orange Balls 6" Diameter	$15.00	$30.00
Total Quantity	23					
					Total Cost	$478.00

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.



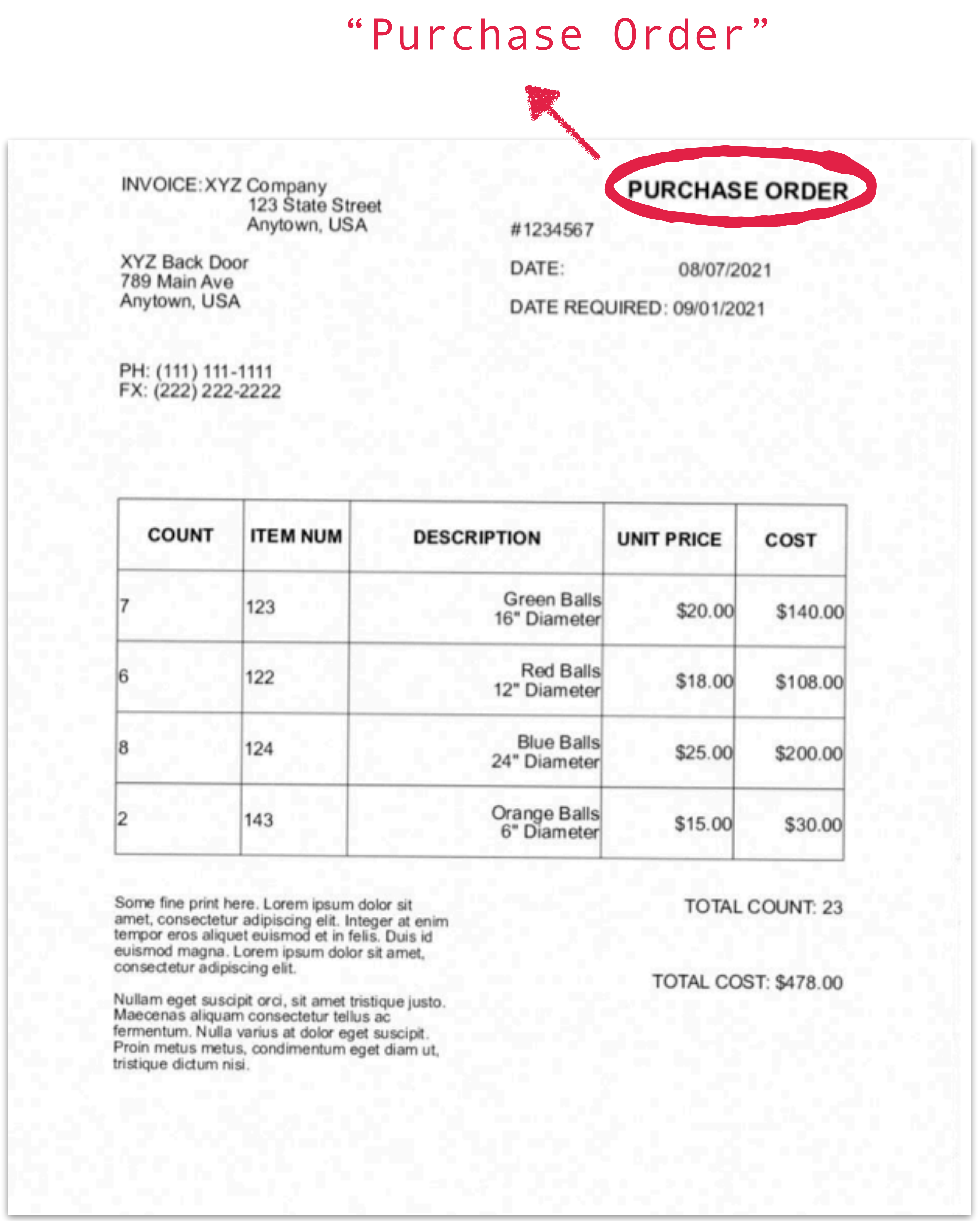
[Processing Data]



[Processing Data]

DIGITIZE & STANDARDIZE DATA

- Shallow pass data collection
- Enough to determine type
- Is there a title like…?
  - “Purchase Order”
  - “Invoice”
  - “Mortgage Application”



“Purchase Order”

INVOICE: XYZ Company
123 State Street
Anytown, USA

PURCHASE ORDER

#1234567

XYZ Back Door
789 Main Ave
Anytown, USA

DATE: 08/07/2021

DATE REQUIRED: 09/01/2021

PH: (111) 111-1111
FX: (222) 222-2222

COUNT	ITEM NUM	DESCRIPTION	UNIT PRICE	COST
7	123	Green Balls 16" Diameter	$20.00	$140.00
6	122	Red Balls 12" Diameter	$18.00	$108.00
8	124	Blue Balls 24" Diameter	$25.00	$200.00
2	143	Orange Balls 6" Diameter	$15.00	$30.00

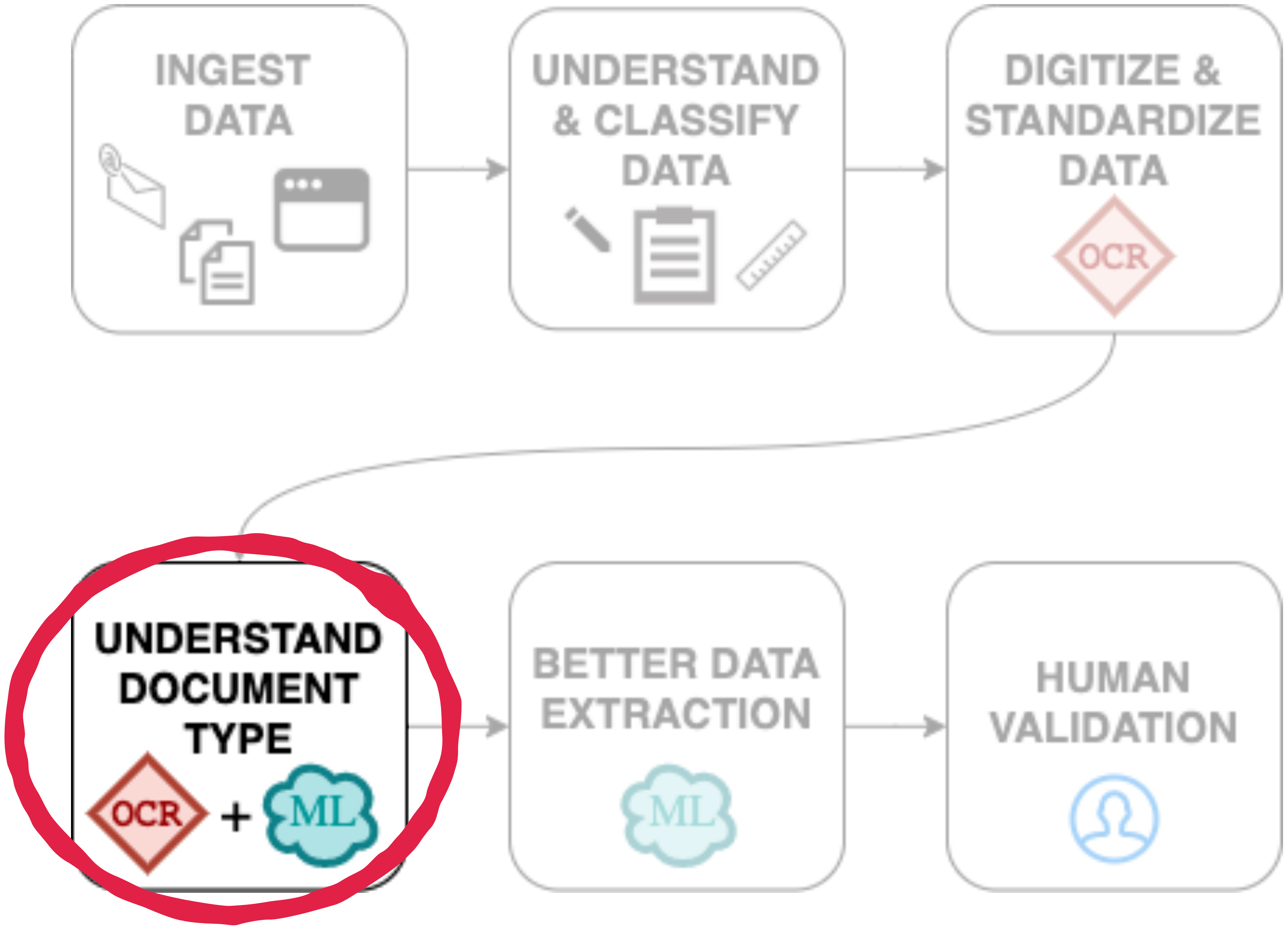
Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

TOTAL COUNT: 23

TOTAL COST: $478.00

[Processing Data]

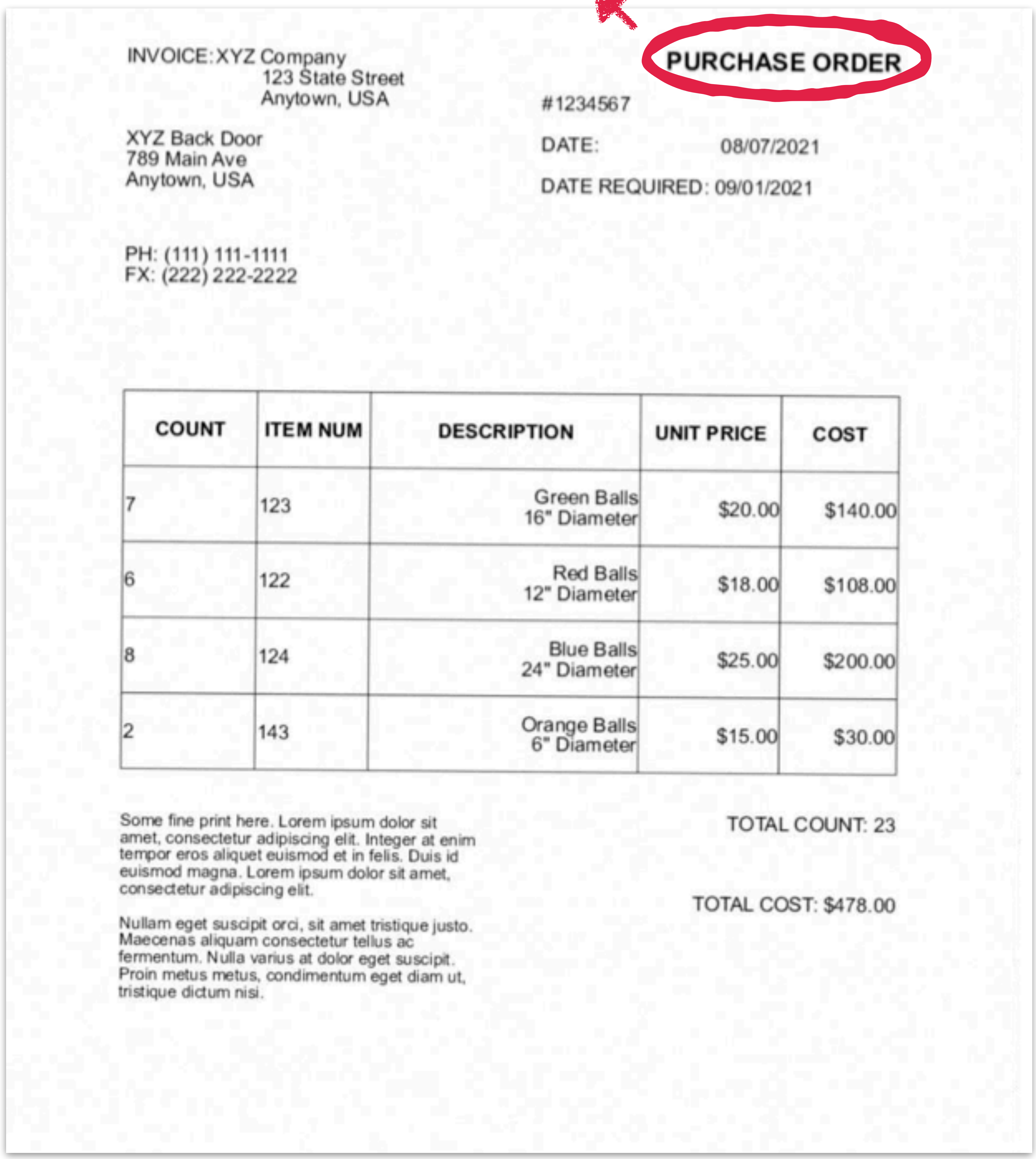


[Processing Data]

UNDERSTAND DOCUMENT TYPE

- Structured data is easy to understand type
- Semistructured & unstructured can be more difficult
- Are more than one document type ingested into the same system?
  - Purchase Orders
  - Invoices
  - Mortgage Applications

DOCUMENT TYPE:
PURCHASE ORDER



INVOICE:XYZ Company
123 State Street
Anytown, USA

PURCHASE ORDER

#1234567

XYZ Back Door
789 Main Ave
Anytown, USA

DATE: 08/07/2021

DATE REQUIRED: 09/01/2021

PH: (111) 111-1111
FX: (222) 222-2222

COUNT	ITEM NUM	DESCRIPTION	UNIT PRICE	COST
7	123	Green Balls 16" Diameter	$20.00	$140.00
6	122	Red Balls 12" Diameter	$18.00	$108.00
8	124	Blue Balls 24" Diameter	$25.00	$200.00
2	143	Orange Balls 6" Diameter	$15.00	$30.00

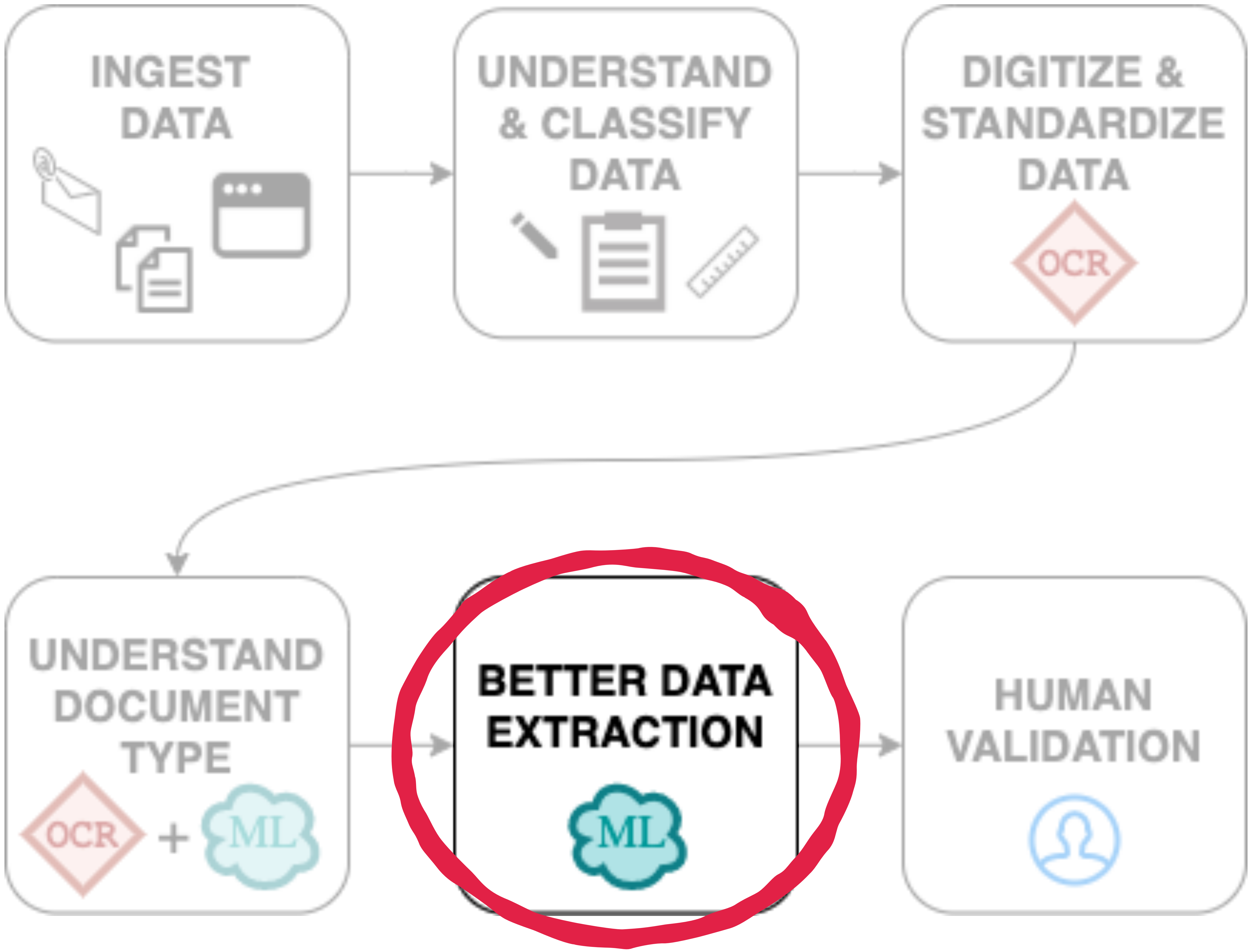
Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

TOTAL COUNT: 23

TOTAL COST: $478.00

[Processing Data]



[Processing Data]

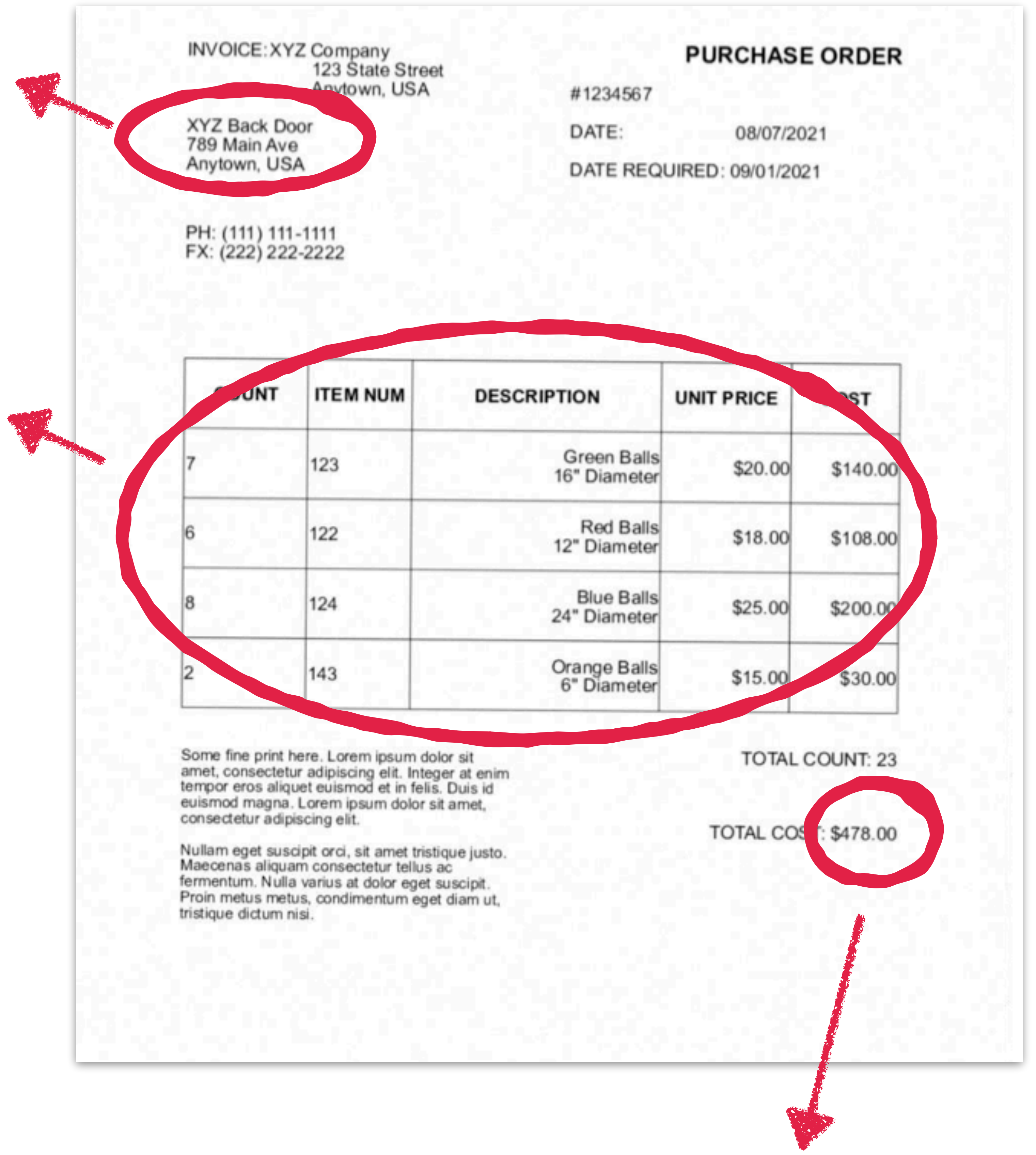
BETTER DATA EXTRACTION

ML

- Knowing the document type, we can extract more specifics
  - Shipping Information
  - Item Details Table
  - Total Cost
- If classified as semistructured or unstructured, we may need to leverage machine learning to map values to the correct keys.

Shipping Address

Item Details



INVOICE: XYZ Company
123 State Street
Anytown, USA

PURCHASE ORDER

XYZ Back Door
789 Main Ave
Anytown, USA

#1234567

DATE: 08/07/2021

DATE REQUIRED: 09/01/2021

PH: (111) 111-1111
FX: (222) 222-2222

COUNT	ITEM NUM	DESCRIPTION	UNIT PRICE	COST
7	123	Green Balls 16" Diameter	$20.00	$140.00
6	122	Red Balls 12" Diameter	$18.00	$108.00
8	124	Blue Balls 24" Diameter	$25.00	$200.00
2	143	Orange Balls 6" Diameter	$15.00	$30.00

Some fine print here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer at enim tempor eros aliquet euismod et in felis. Duis id euismod magna. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

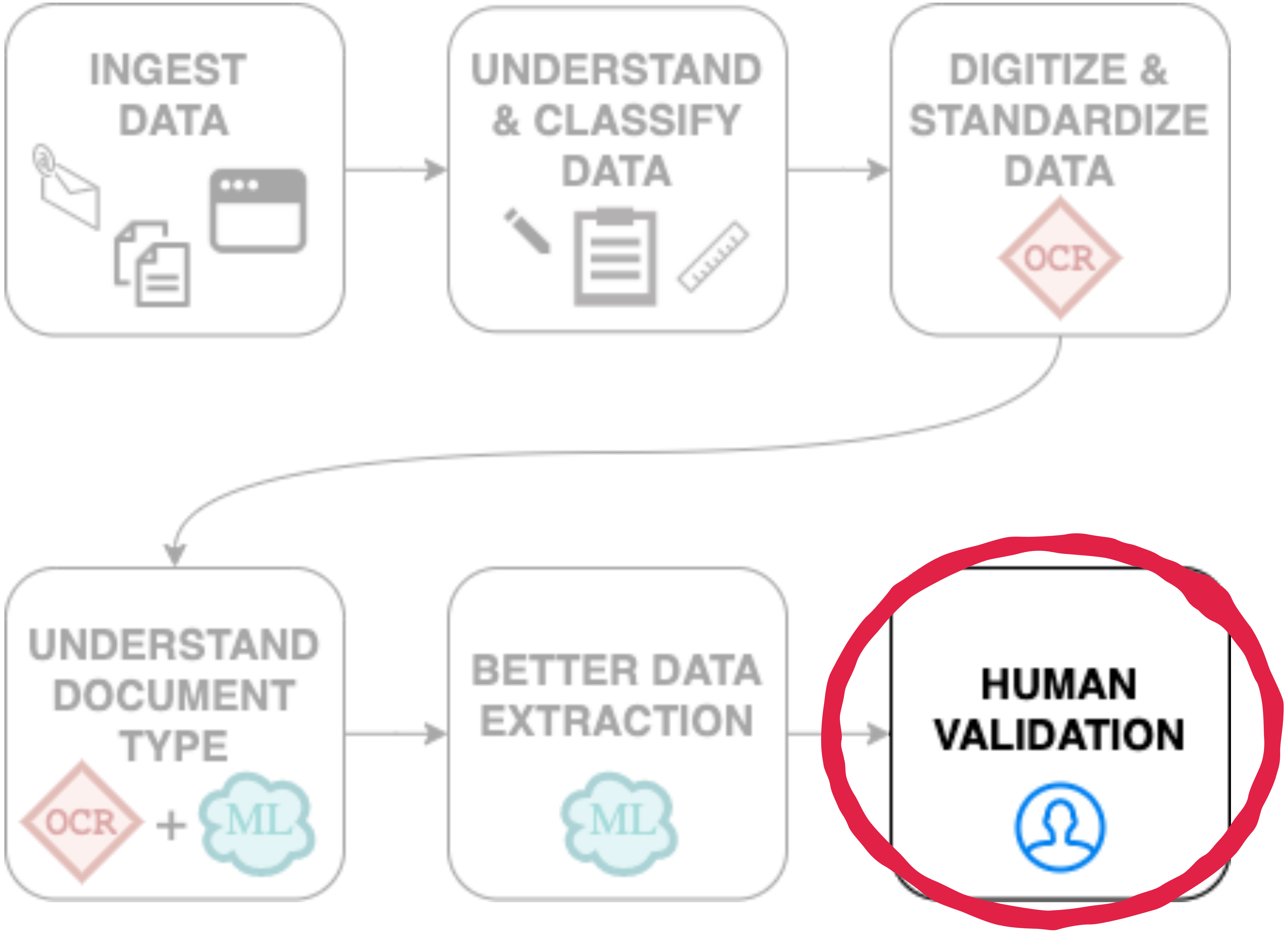
Nullam eget suscipit orci, sit amet tristique justo. Maecenas aliquam consectetur tellus ac fermentum. Nulla varius at dolor eget suscipit. Proin metus metus, condimentum eget diam ut, tristique dictum nisi.

TOTAL COUNT: 23

TOTAL COST: $478.00

Total Cost

[Processing Data]



[Processing Data]

HUMAN VALIDATION

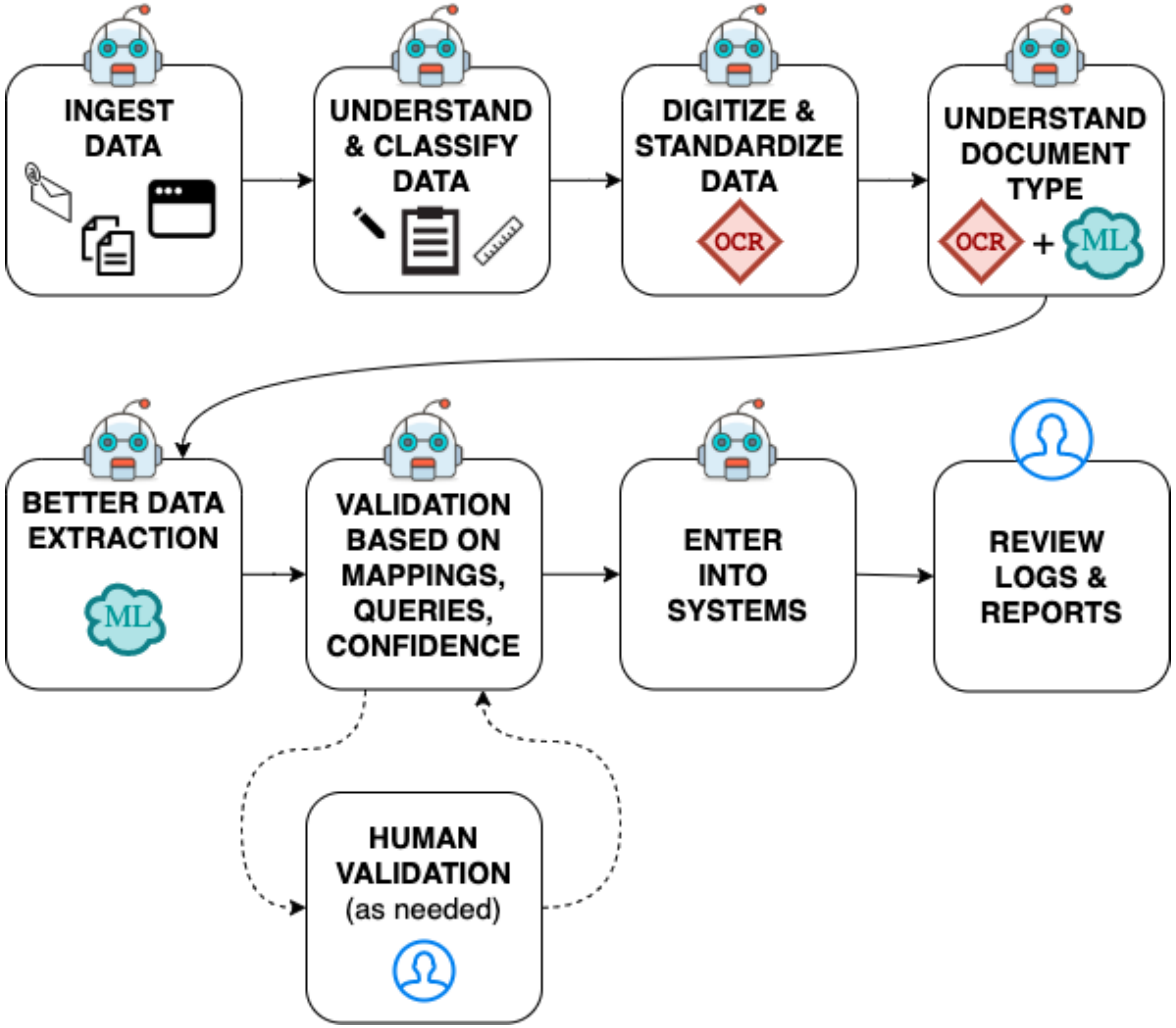
- Certain level of human-in-the-loop is necessary (and good)
  - Confirm accurate information
- Machine learning models improve with more samples and verifications
- Some cross-checking can be automated

How do we link this all together?

[Single Pipeline]

Robotic Process Automation

- Also known as RPA
- Method of automating repetitive tasks
- Often use the screen same as a live user would
  - Avoids the wait/requirement for application integration
- Perform faster, with accuracy, can run without stopping


INGEST DATA
UNDERSTAND & CLASSIFY DATA
DIGITIZE & STANDARDIZE DATA
OCR
UNDERSTAND DOCUMENT TYPE
OCR + ML
BETTER DATA EXTRACTION
ML
VALIDATION BASED ON MAPPINGS, QUERIES, CONFIDENCE
ENTER INTO SYSTEMS
REVIEW LOGS & REPORTS
HUMAN VALIDATION (as needed)

[Single Pipeline]

Case Study: PepsiCo

PEPSICO

- CHALLENGE
  - Slow document processing
  - Prone to errors
  - Very manual

- SOLUTION
  - Single process flow
  - ABBYY FlexiCapture

- RESULTS
  - Lower error rates
  - Reduced processing times
  - Team worked more efficiently

PepsiCo automates invoice processing with ABBYY FlexiCapture, 2021. [Online; accessed 13-August2021].

[Single Pipeline]

Case Study: Sysco

- CHALLENGE
  - Very manual processes
  - Prone to errors
  - Not enough time

- SOLUTION
  - Robotic Process Automation
  - Multiple departments and processes

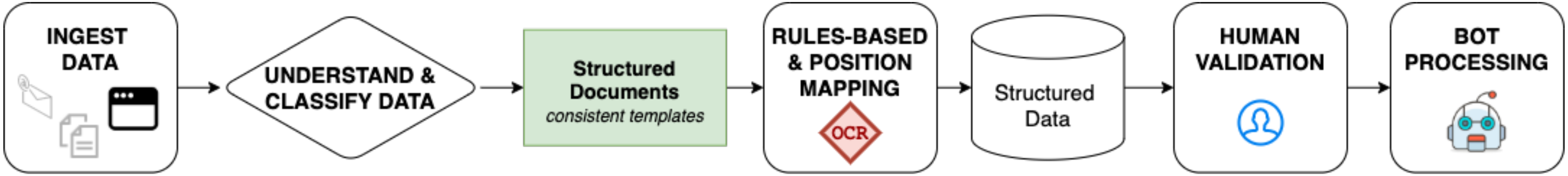
- RESULTS
  - Over 200,000 hours productivity saved in first 3.5 years
  - Built a culture of automation

Alison Major and Kim Meredith. Sysco: Scaling an rpa program without compromise, 2021. [Online; accessed 15-August2021].

Blue Prism Cafe´. Sysco: Scaling an intelligent automation program without compromise, 2021. [Webinar; accessed 15- August2021].

[Single Pipeline]

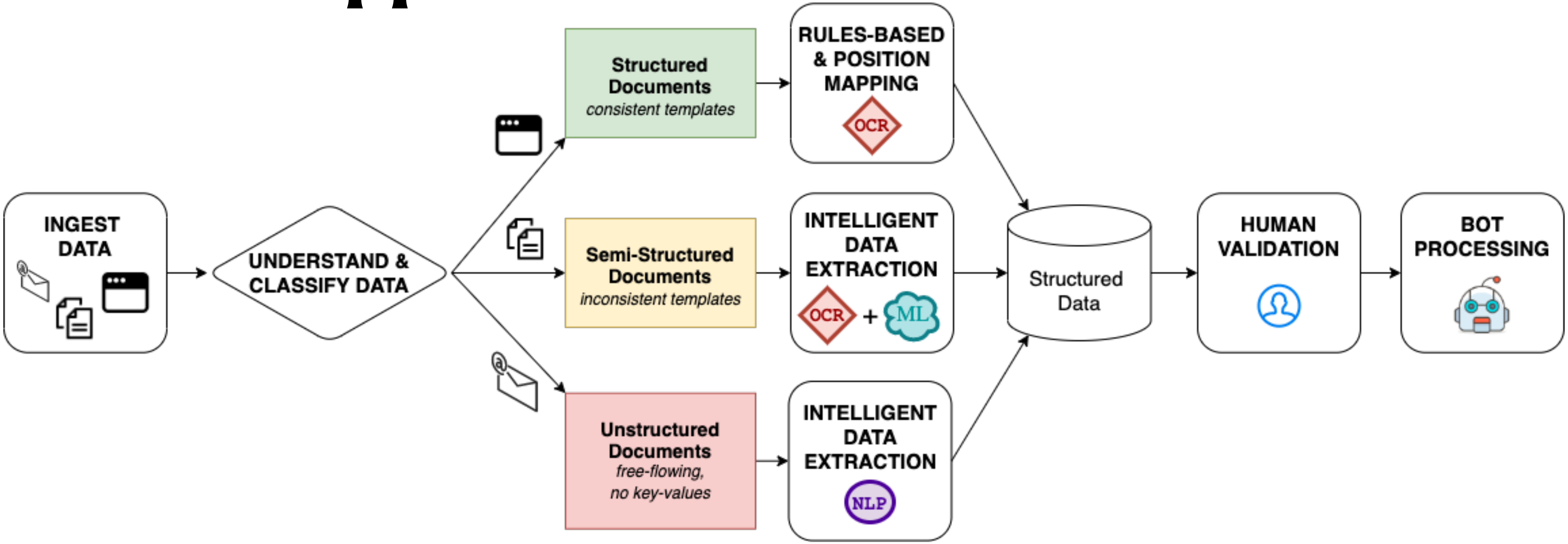
General Application



- Start with a single input method
- One classification (structured data)
- One document type
- Automate end-to-end using RPA

[Single Pipeline]

General Application


INGEST DATA
UNDERSTAND & CLASSIFY DATA
Structured Documents
consistent templates
Semi-Structured Documents
inconsistent templates
Unstructured Documents
free-flowing, no key-values
RULES-BASED & POSITION MAPPING
OCR
INTELLIGENT DATA EXTRACTION
OCR + ML
INTELLIGENT DATA EXTRACTION
NLP
Structured Data
HUMAN VALIDATION
BOT PROCESSING

- Begin adding new ingestion methods that automatically tie into the existing process

[Solutions & Costs]

ABBYY FlexiCapture

- Best NLP, ML
- Starting at $0.028 per page plus training/setup costs

Amazon Textract

- Powerful OCR
- Built-in Human-in-the-Loop
- $0.0015 per page

Google Tesseract

- Open Source
- Less accurate
- Free

Microsoft PowerAutomate

- Variety of solutions available
- Easy to link processes

Monarch

- Good for Structured & Unstructured Data
- Desktop based

Blue Prism Decipher

- RPA Software
- Direct integration to OCR with human-in-the-loop

[Conclusion]

Is there a single method to receive multiple forms of information and automatically pull the desired data?

Many tools are available, but they can be combined into a single solution.

Consider including RPA to make your solution more robust.

Start small, then build more methods to ingest different document types.