

How to Convert PDF to Excel Spreadsheet:

A Practical Workflow



- PDFs are built for presentation; Excel needs structured, reusable data



- Goal: repeatable workflow for audit-friendly, accurate spreadsheets



- Suitable for operations, research, finance-support, and admin tasks



- End-to-end process: inspect, choose method, convert, clean, validate



Why PDF-to-Excel Conversion Breaks



- PDFs store characters at fixed coordinates—not rows or columns
- No real cells exist, so converters guess table structure
- Common failures: split columns, merged cells, repeated headers
- Wrapped text can create extra rows; page breaks split tables
- Scanned PDFs need OCR—a fundamentally different process

Match the Method to the PDF Type



- Classify first: selectable text or image-only scan



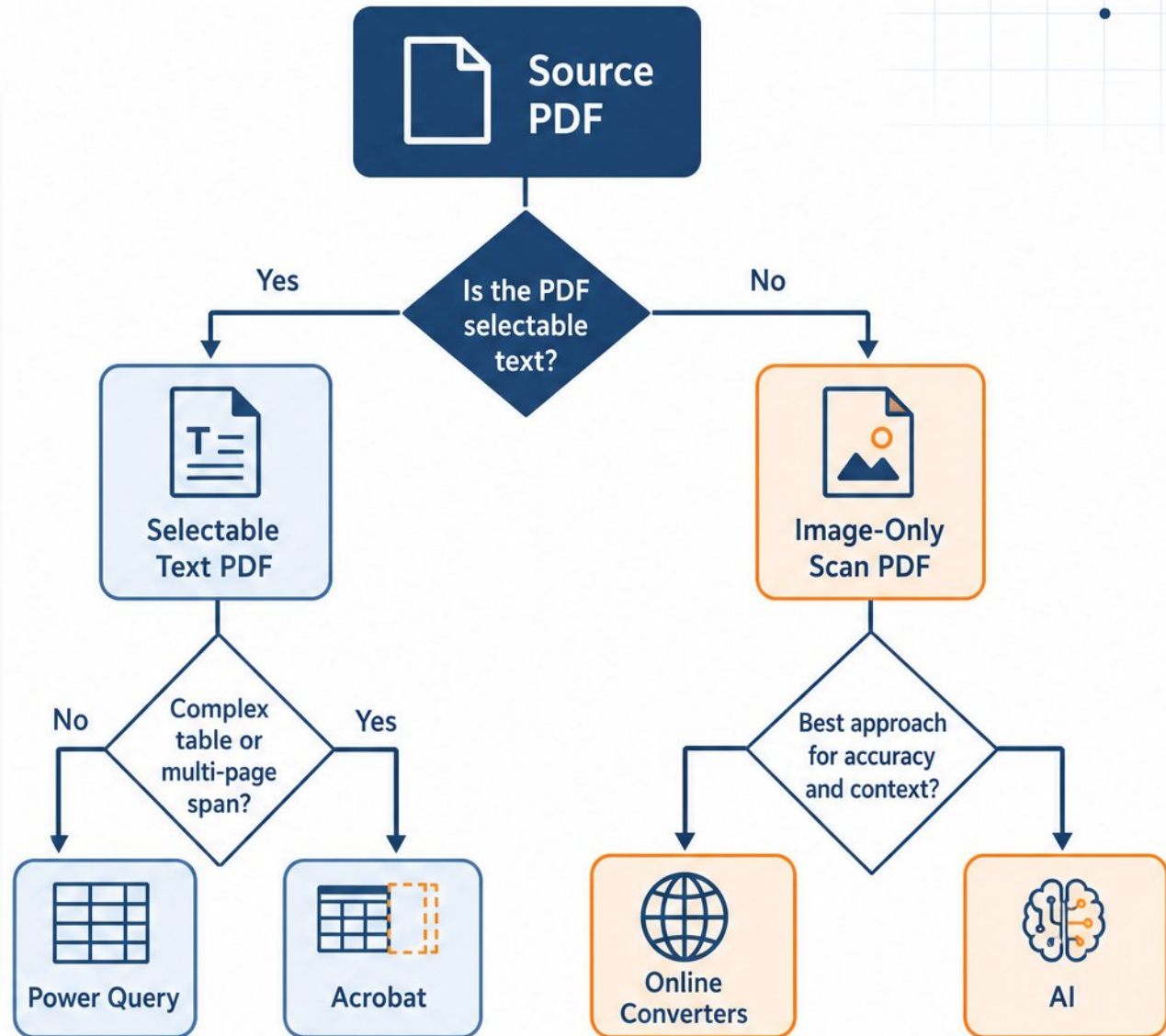
- Consider table complexity and page span



- Compare tools: Power Query, Acrobat, online converters, AI



- Weigh file type, volume, sensitivity, and frequency





Best First Option: Excel's Built-In PDF Connector

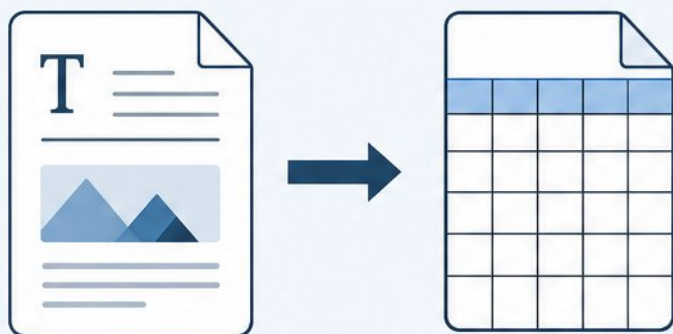
-  - Built-in in Excel 2021 or Microsoft 365 on Windows
-  - Free, no extra tools needed for clear, text-based PDFs
-  - Use Data > Get Data > From File > From PDF
-  - Preview detected tables in Navigator before loading
-  - No Mac support; struggles with scans and complex layouts



When to Use Adobe Acrobat or an Online Converter

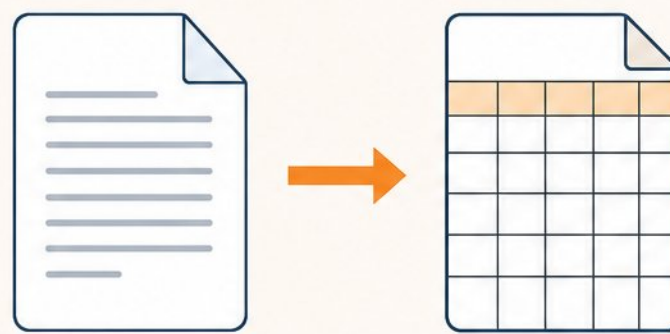


Adobe Acrobat Export



VS

Online Converter



- Use Acrobat export for mixed text and scanned PDFs



- Online converters suit quick, non-sensitive jobs



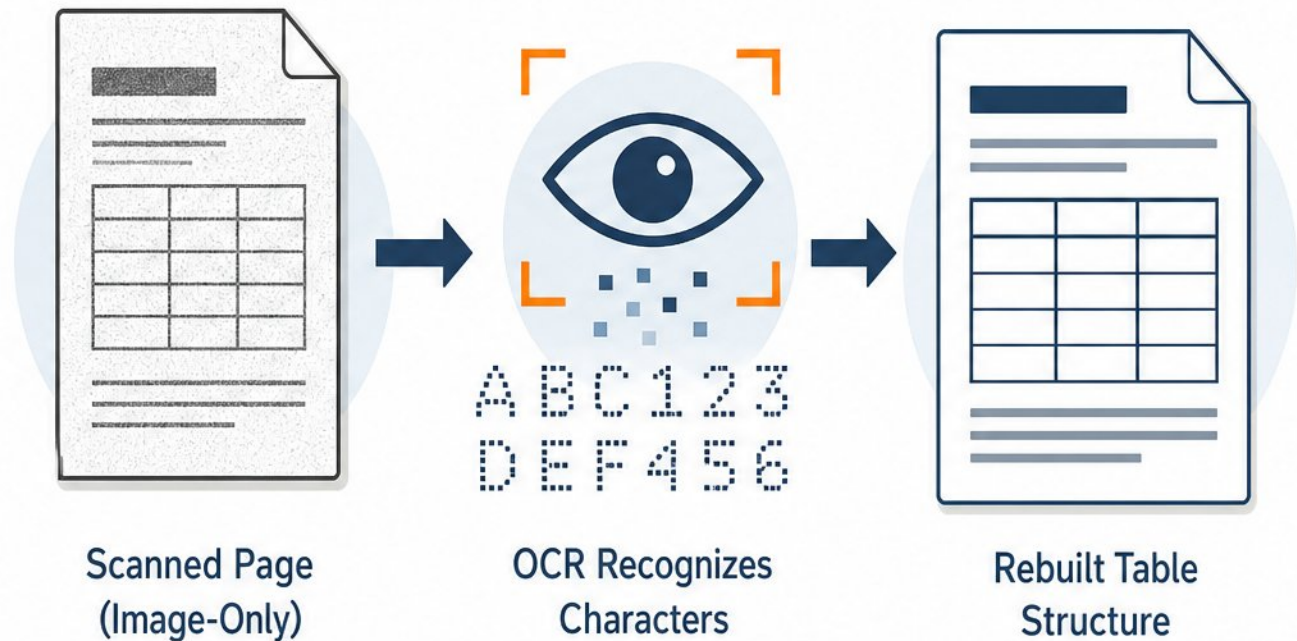
- Avoid them for sensitive or compliance-bound data



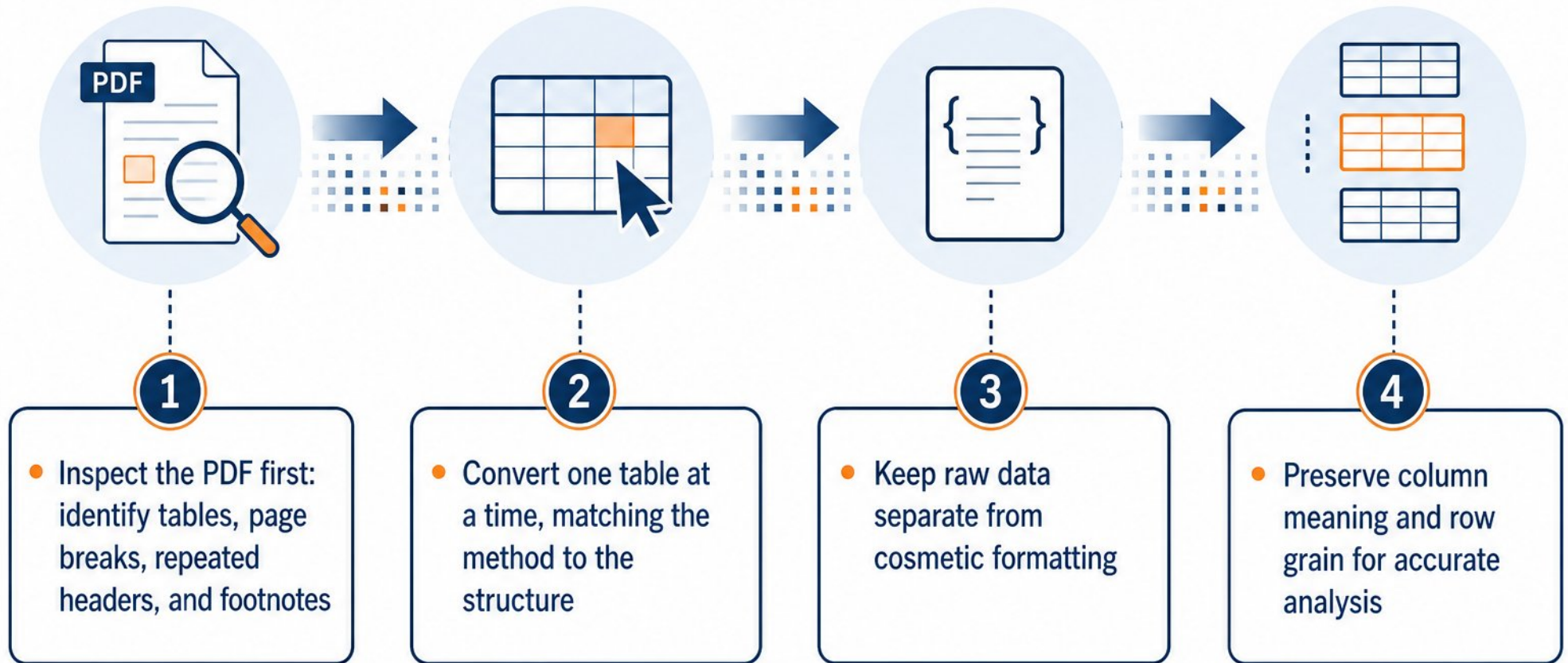
- Compare speed, accuracy, cost, and control first

Handling Scanned PDFs and OCR

- Digital PDFs contain selectable text; scans are image-only
- Test by selecting text to classify the source
- OCR recognizes characters, then rebuilds table structure
- Expect 70-99% accuracy depending on scan quality
- Poor scans or handwritten data: manual keying may be safer

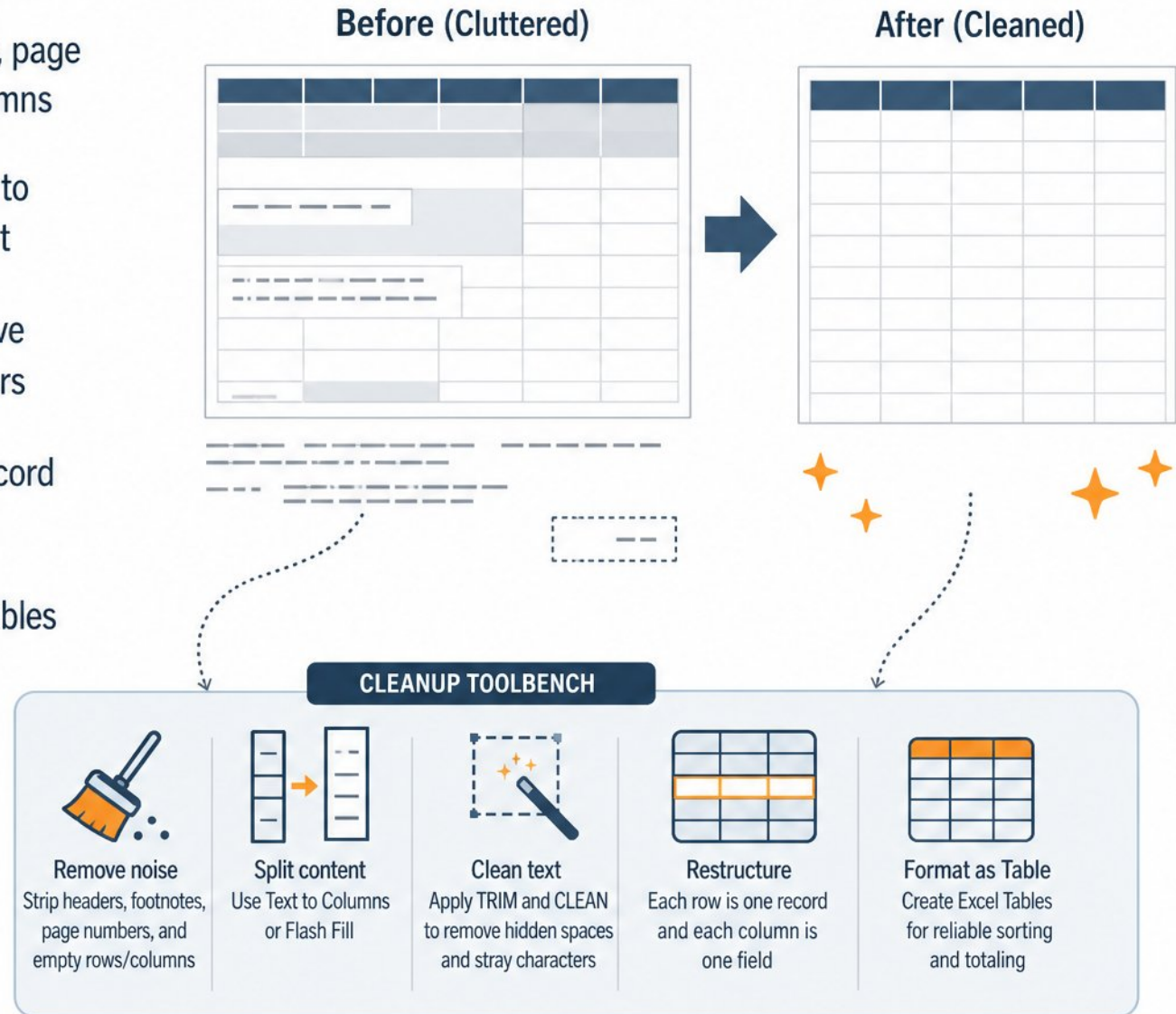


Step-by-Step Workflow for Accurate Conversion



Cleaning and Restructuring the Converted Data

- Strip repeated headers, footnotes, page numbers, and empty rows or columns
- Use Text to Columns or Flash Fill to split merged and wrapped content
- Apply TRIM and CLEAN to remove hidden spaces and stray characters
- Restructure so each row is one record and each column is one field
- Format cleaned ranges as Excel Tables for reliable sorting and totaling



Fixing the Most Common Conversion Errors



- Convert text-formatted numbers and dates to real values via Convert to Number



- Repair merged cells, split rows, and misaligned columns



- Remove stray characters with Find and Replace: \$, commas, parentheses, spaces



- Watch for hidden characters and non-printing artifacts using TRIM and CLEAN

Text		
Text		
Text		

Text-formatted numbers and dates

Merged cells, split rows, and misaligned columns

\$ Text		
Text ,		
(Text)		

Stray characters: \$, commas, parentheses, spaces

Text		
Text		
Text		

Hidden characters and non-printing artifacts

Validating Converted Data for Accuracy



Run three quick checks:
row count, column alignment,
summed totals



Spot-check first, last, and
page-boundary rows



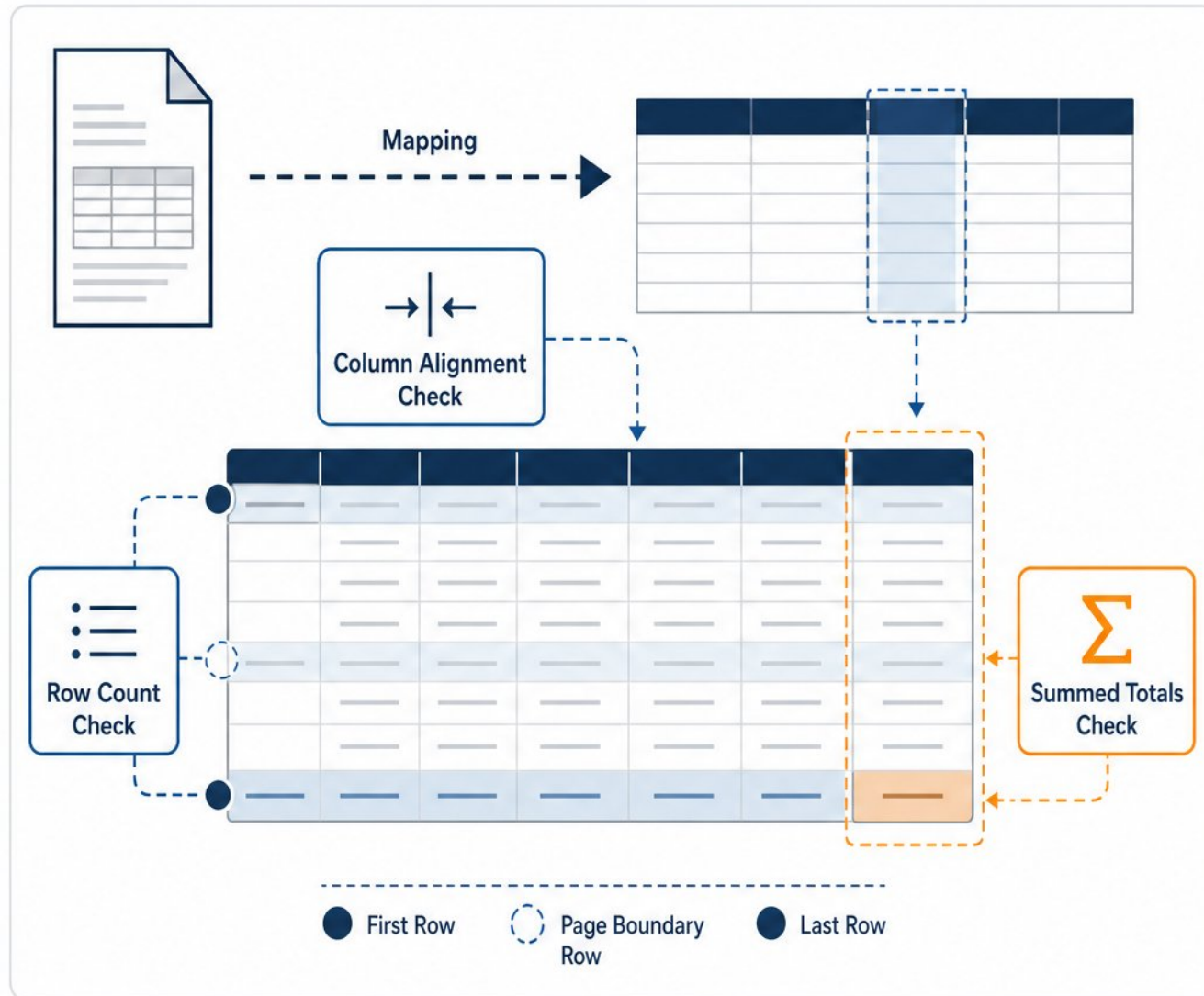
Verify dates, identifiers,
negatives, and
currency values



Use filters, sorting, and
conditional formatting to
surface outliers



Accept only when figures
tie to the source PDF



Working with Complex Financial and Research Tables



- Remove repeated page headers



- Check transitions across page breaks



- Separate summary totals from raw data



- Preserve parentheses, currency, and leading zeros



- Confirm date conventions are consistent

Financial and Research Summary (Continued)						
Category	Current Period		Prior Period		Variance	
	Budget	Actual	Budget	Actual	Amount	%
Research Expenses						
- Personnel	-----	-----	-----	-----	-----	-----
- Materials	-----	-----	-----	-----	-----	-----
- Travel	-----	-----	-----	-----	-----	-----
- Other	-----	-----	-----	-----	-----	-----
Subtotal – Research Expenses	-----	-----	-----	-----	-----	-----
Summary Totals	-----	-----	-----	-----	-----	-----

↓ Page Break ↓

(Continued)						
Category	Current Period		Prior Period		Variance	
	Budget	Actual	Budget	Actual	Amount	%
Other Operating Expenses						
- Item A	-----	-----	-----	-----	-----	-----
- Item B	-----	-----	-----	-----	-----	-----
- Item C	-----	-----	-----	-----	-----	-----
Subtotal – Other Operating Expenses	-----	-----	-----	-----	-----	-----
Summary Totals	-----	-----	-----	-----	-----	-----

Automation and Repeatable Processes



- Identify recurring PDF formats eligible for Power Query import routines



- Save cleanup steps as reusable Power Query transformations



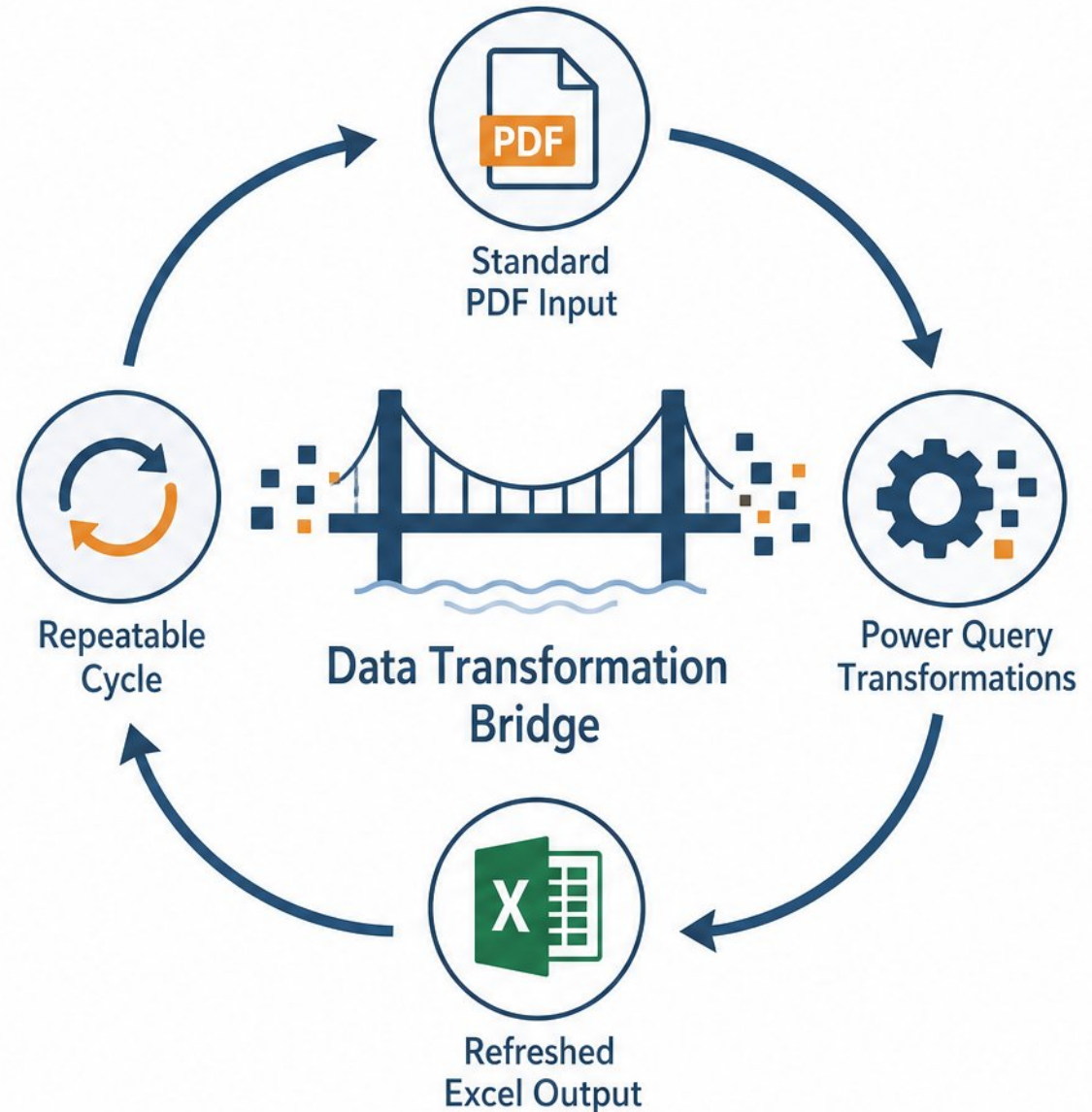
- Refresh results automatically when a new PDF version arrives



- Balance automation with manual review for financial data



- Guard against layout changes and traceability issues



Practical Checklist and Common Pitfalls



- Classify the PDF, convert, then clean and validate output



- Verify row counts, totals, and key figures against source



- Watch for split rows, merged cells, and misread dates

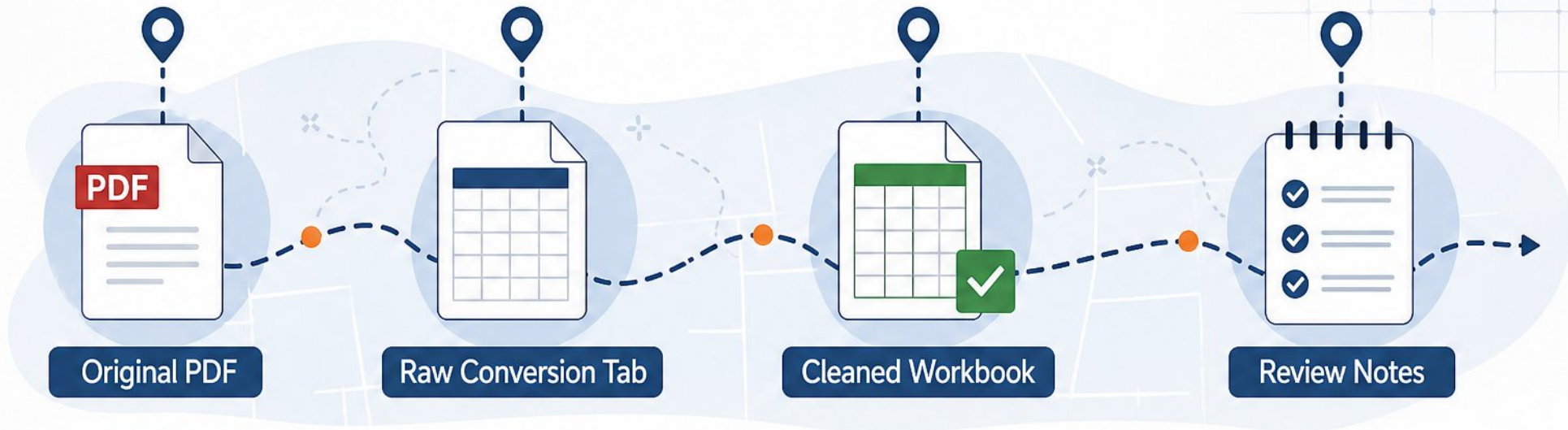


- Look for overlooked footnotes and text-formatted numbers



- Acceptance: figures tie to the source, ready for use

Next Steps for Better Data Quality



- ✓ Keep original PDF and raw conversion tab for traceability
- ✓ Apply checklist consistently to recurring reports
- ✓ Build a library of standard clean-up transformations
- ✓ Escalate if quality is insufficient: request original source
- ✓ Turn to dedicated OCR tools or consult a data specialist

PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/0598026f/public