WIPO REGIONAL TRAINING WORKSHOP ON IP DATA VALIDATION, VERIFICATION AND EXCHANGE

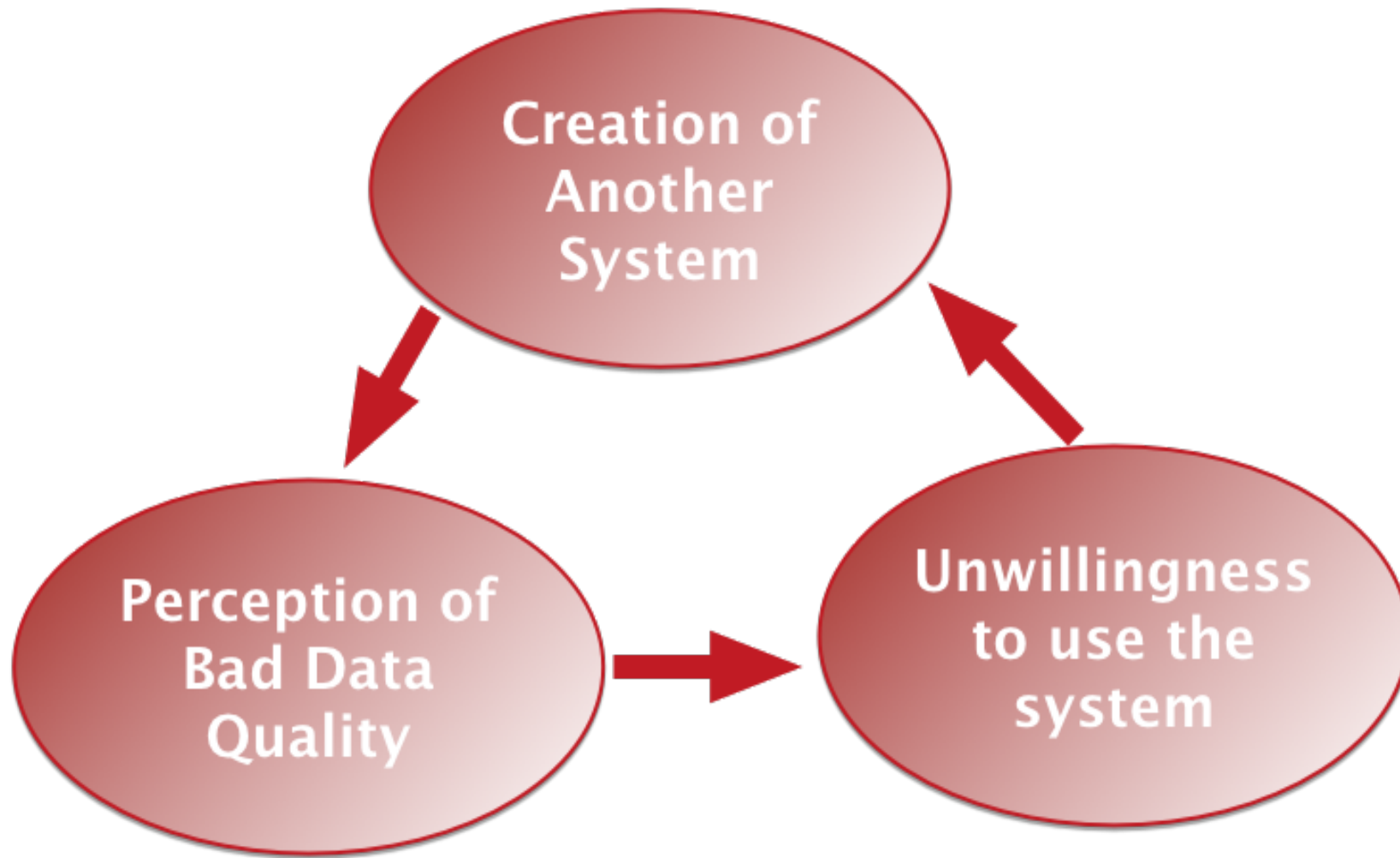## DATA VALIDATION AND VERIFICATION USING IPOBSD'S TOOLS

Gregory Sadyalunda
Project Manager
IP Office Business Solutions Division (IPOBSD)
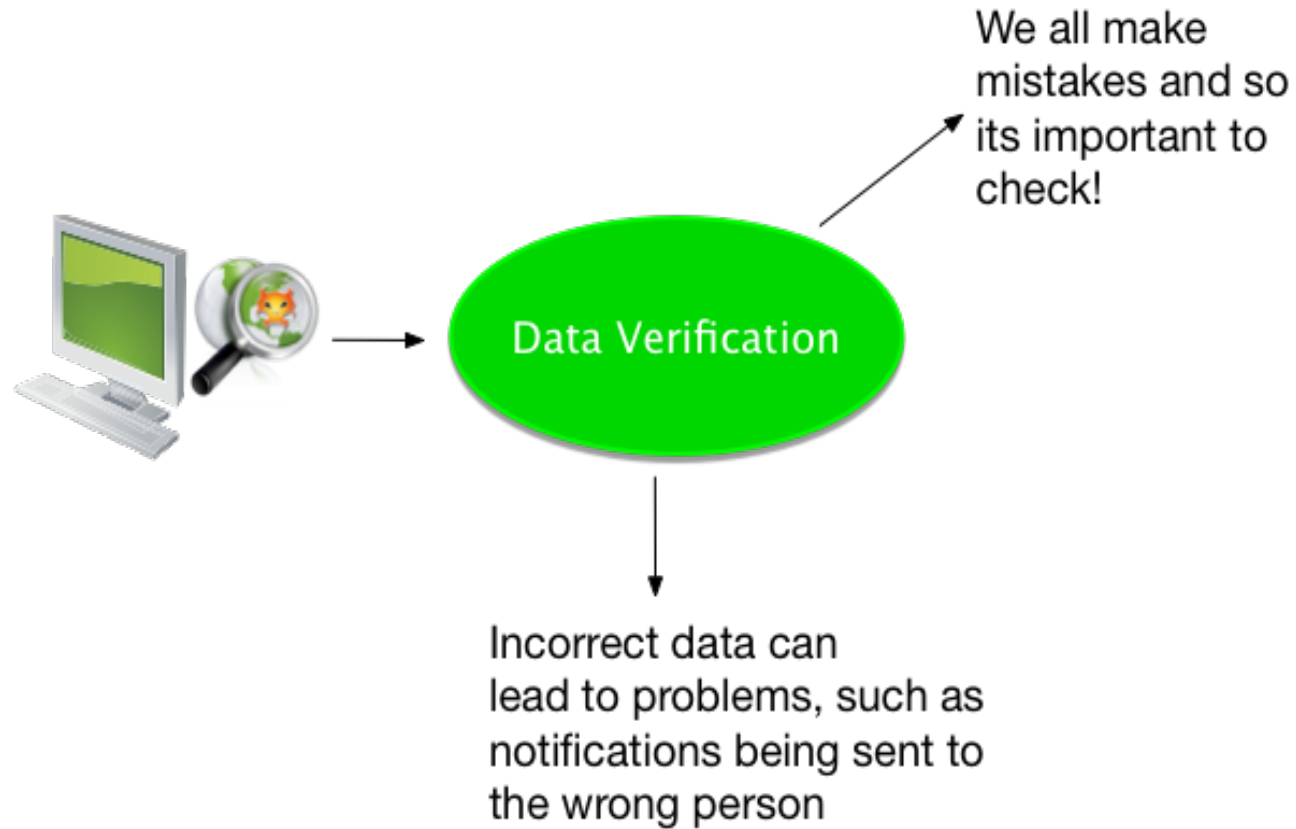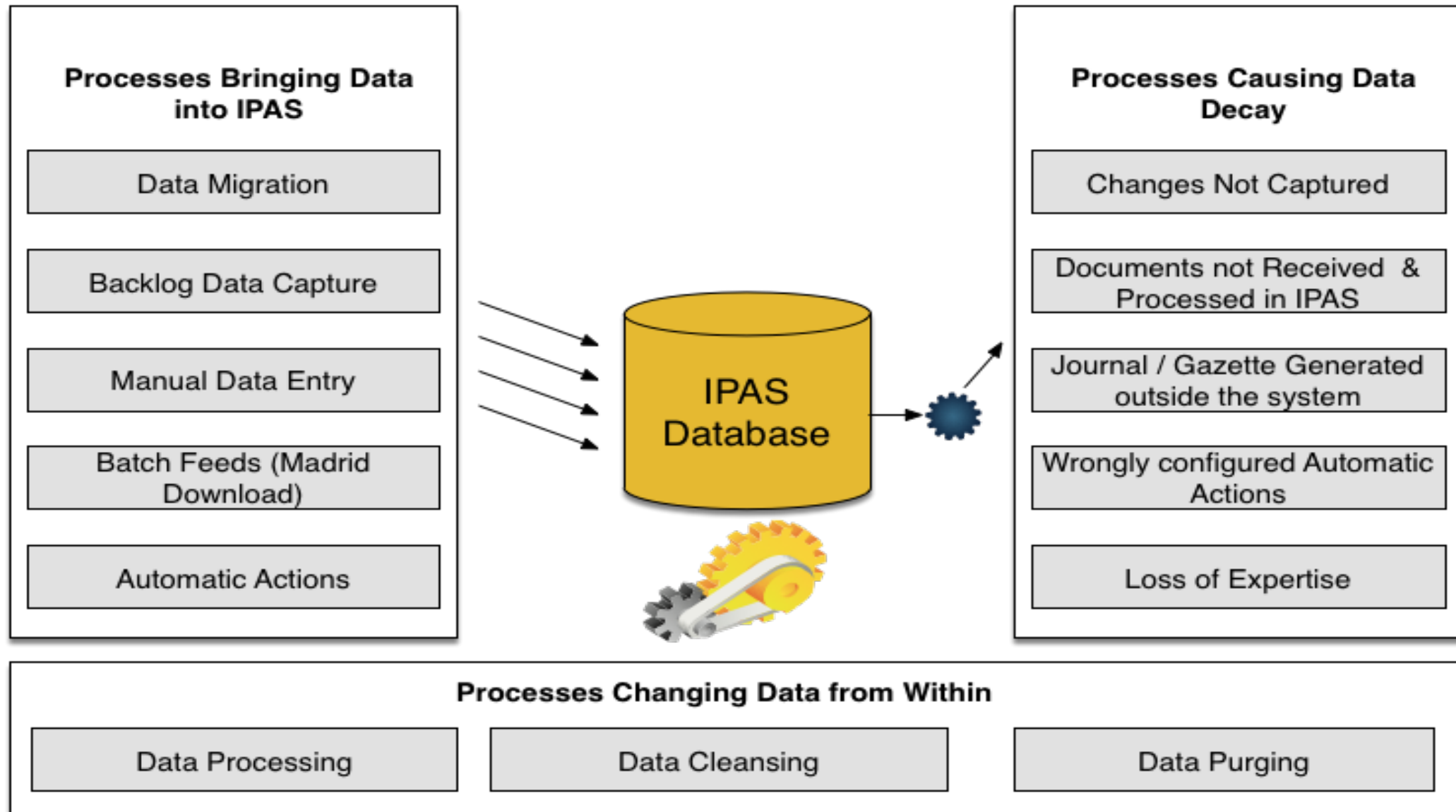
Data Quality Vicious Cycle

What is Data Verification?

- Check for spelling mistakes

- Check for data that has not been entered correctly

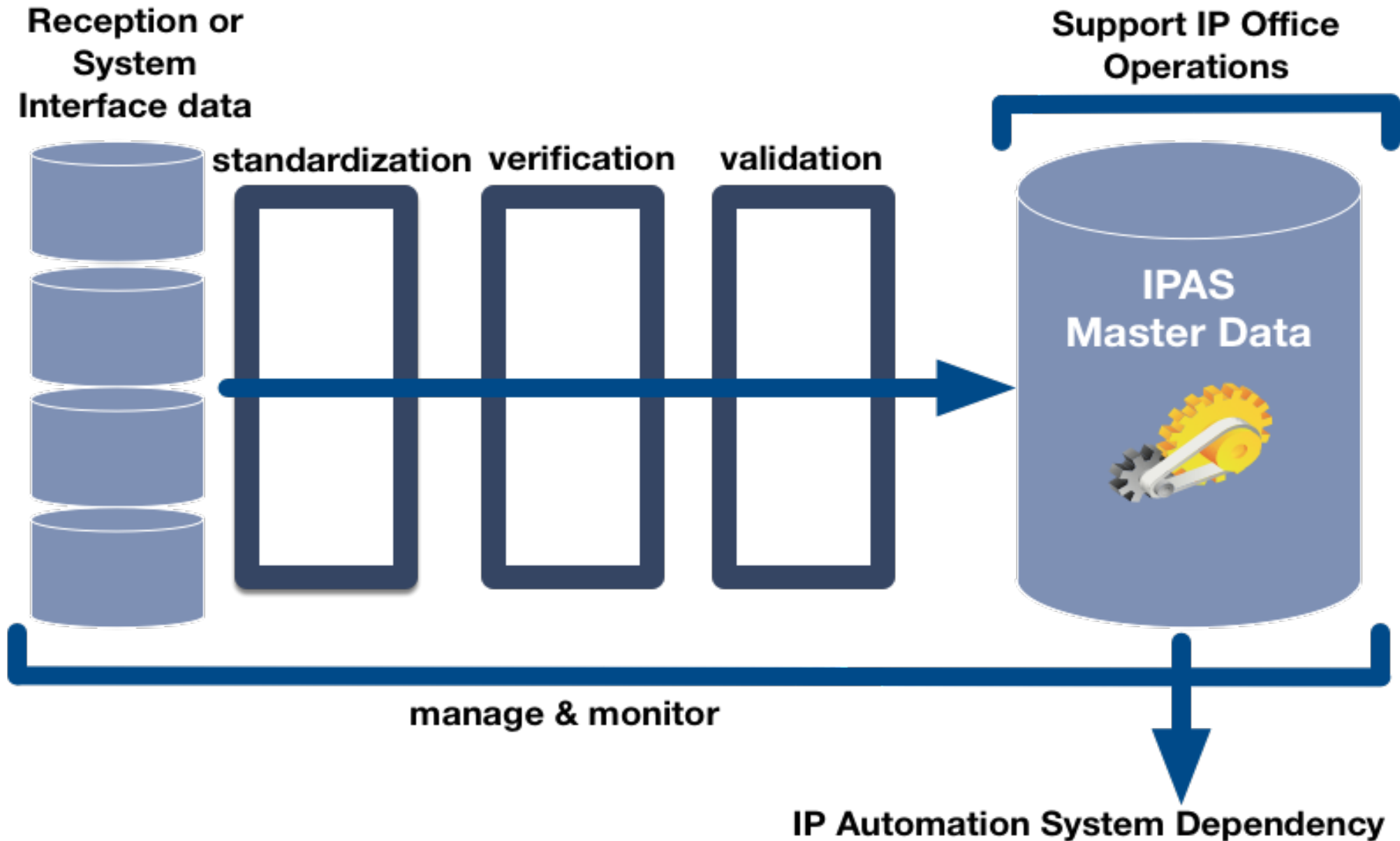- Check that all data has been entered and nothing is missing
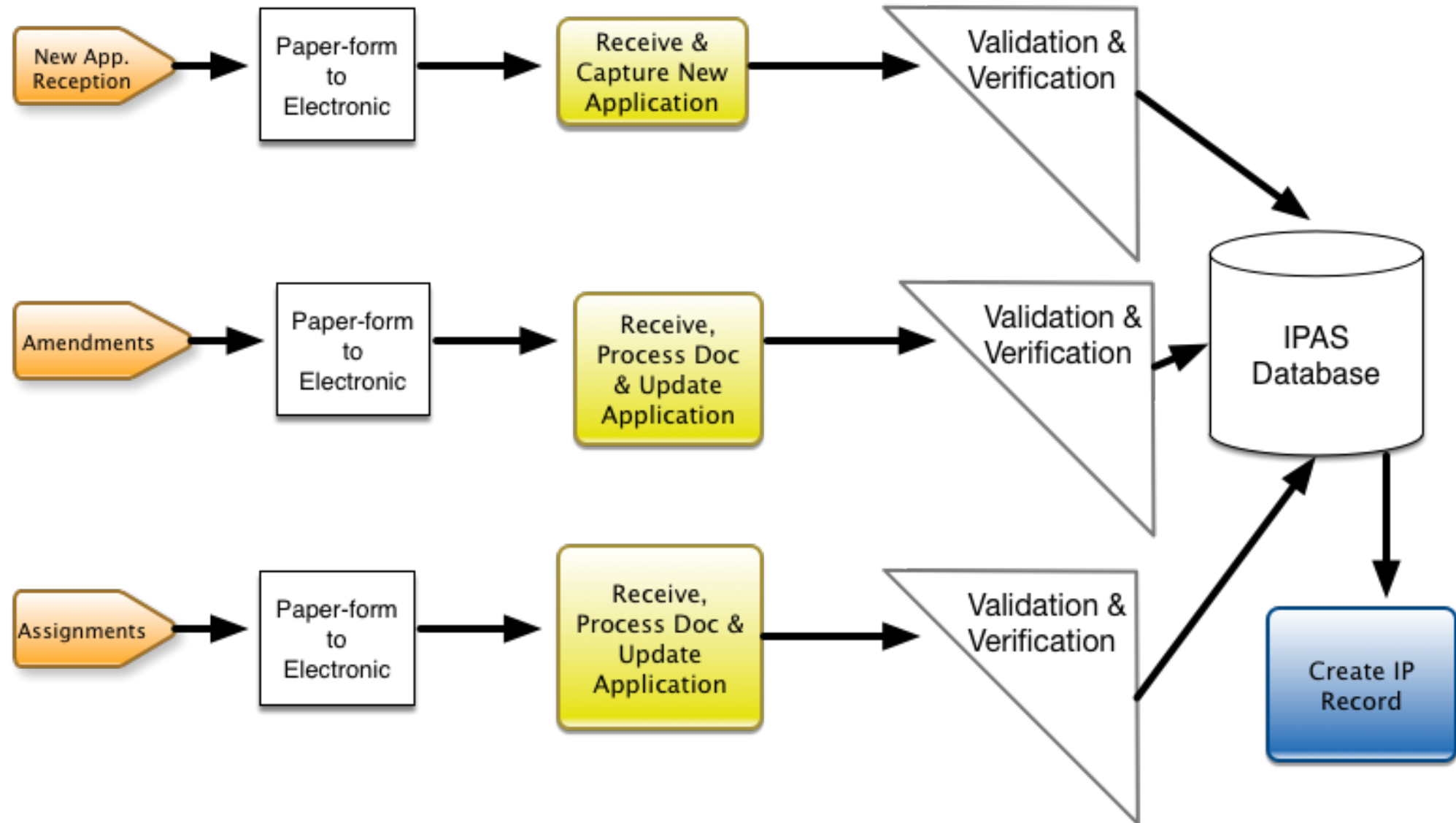
Why do you check the data?

We all make
mistakes and so
its important to
check!

Data Verification

Incorrect data can
lead to problems, such as
notifications being sent to
the wrong person

Causes of Data Quality Problems

Processes Bringing Data into IPAS

Data Migration

Backlog Data Capture

Manual Data Entry

Batch Feeds (Madrid Download)

Automatic Actions

IPAS Database

Processes Causing Data Decay

Changes Not Captured

Documents not Received & Processed in IPAS

Journal / Gazette Generated outside the system

Wrongly configured Automatic Actions

Loss of Expertise

Processes Changing Data from Within

Data Processing

Data Cleansing

Data Purging

Building the Best & Reliable IP Database

Best Practices – IPAS Data Capture

Data Standardization

Unstandardized Data	Standardized Data
Microsoft Corp.	Microsoft Corporation
DataFlux Inc	DataFlux Incorporated
Windmill Co.	Windmill Company
MR RAMSAY SHONGE	Mr. Ramsay Shonge
Mister Ramsay Shonge	Mr. Ramsay Shonge
Mster Robert Chiweza	Mr. Robert Chiweza
#(877)8463589 EXT250	877-846-3589 ext 250
8778463589 x250	877-846-3589 ext 250
Adams and Adams	Adams & Adams
COCA-COLA	Coca Cola
Int'l Paper Company	International Paper Company

Defining Issues with your Data - Sample

Data Sources

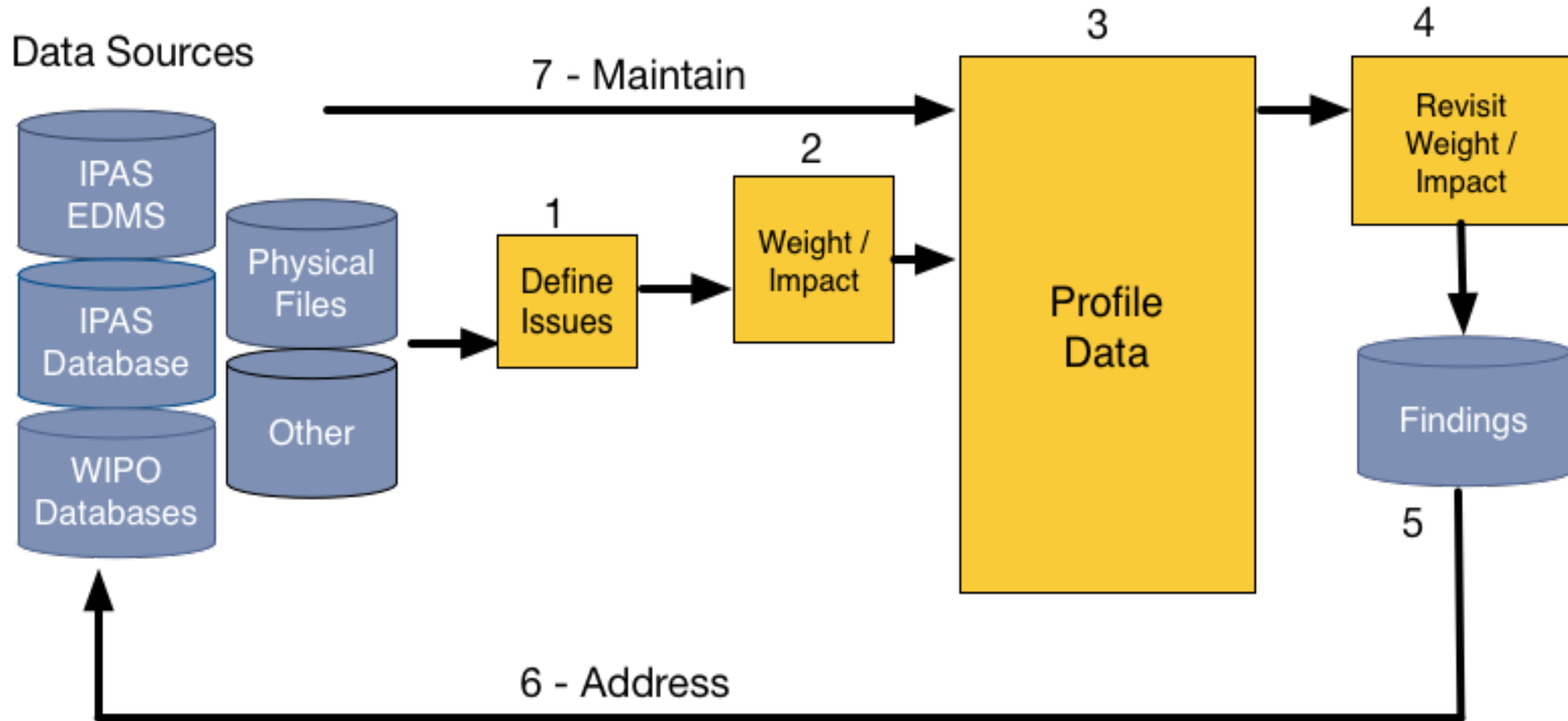| |
|---|
| Potential Duplicates |
| Files in Wrong statuses |
| Wrong Automatic Actions |
| Out of Range Values |
| Changes Not Captured |
| Missing Data Values |
| Wrong Data Values |
| Potential Defaults |
| Journal Processed outside |
| Invalid Data Values |
| Documents done outside |
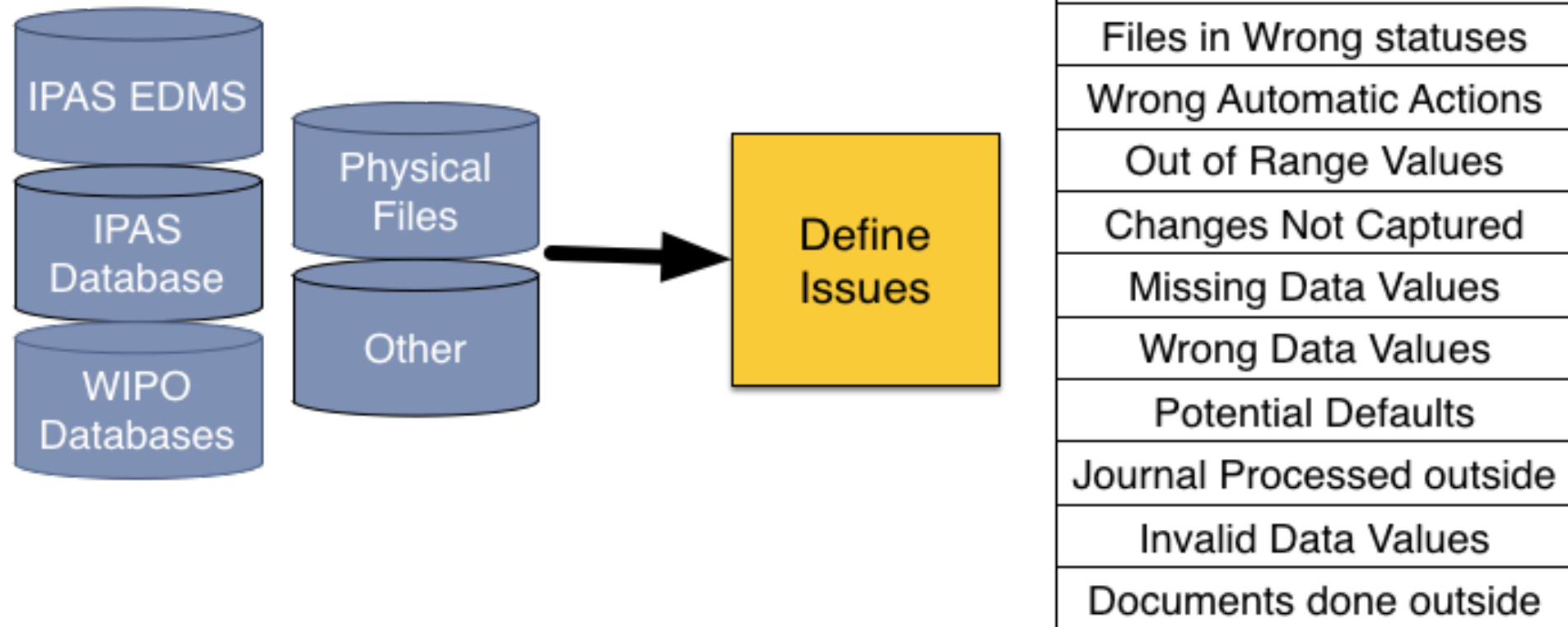
Weight Impact

Data Sources
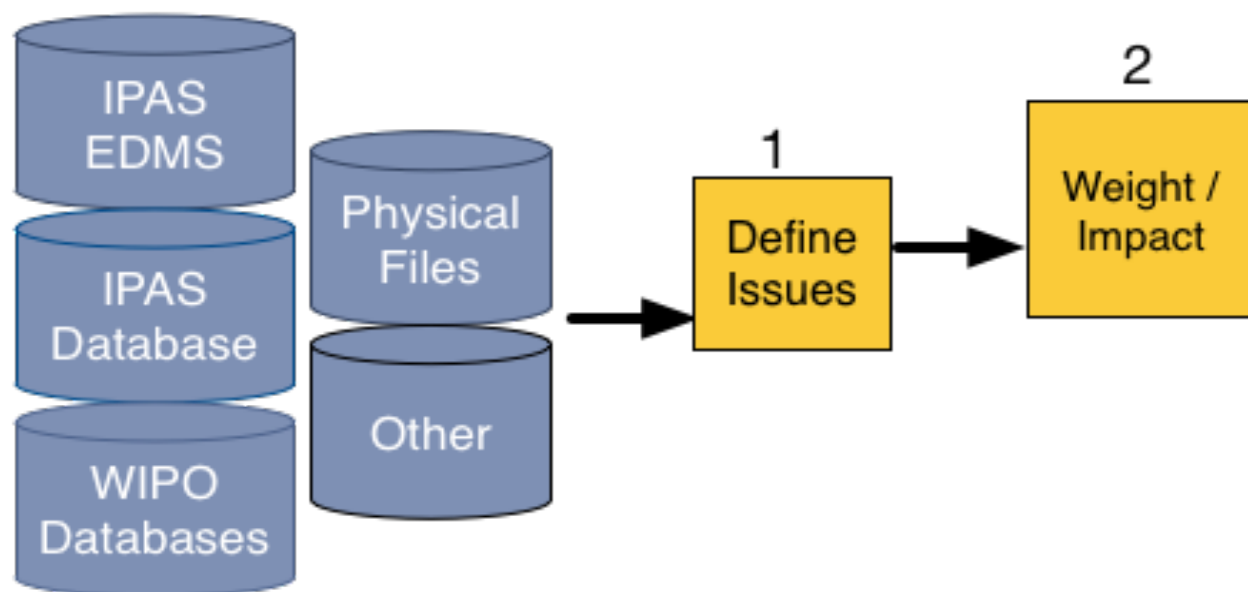
- After the issues are initially identified:
  - Some issues are more critical than others
  - Weights are not priorities
  - Assign a weighting factor (1-5)
  - Weighting factors SHOULD change by project

Assessing Your Data

Data Sources

IPAS EDMS

IPAS Database

WIPO Databases

Physical Files

Other

7 - Maintain

1 Define Issues

2 Weight / Impact

3 Profile Data

4 Revisit Weight / Impact

5 Findings

6 - Address

What is Data Profiling?

- The use of analytical techniques on data for the purpose of developing a thorough knowledge of its content, structure and quality.

- A process of developing information about data instead of information from data.

What is Data Profiling?

Information About Data: (Data Profiling)

- 70% of applicants addresses have no city name
- the range of values in file series is 1900 to 2099
- there are 1500 files with in Migrated Status
- there are 2000 files missing from the system

Information FROM Data: (not Data Profiling)

- Texas auto buyers buy more Cadillacs per capita than any other state
- The average mortgage amount increased last year by 6%
- 10% of last year's customers did not buy anything this year

All data when it is initially entered into the system should be checked for errors so that bad data does not get put onto permanent disk files. Remember the rule: "Garbage in, garbage out!" This process of error checking is called VALIDATING OR EDITING.
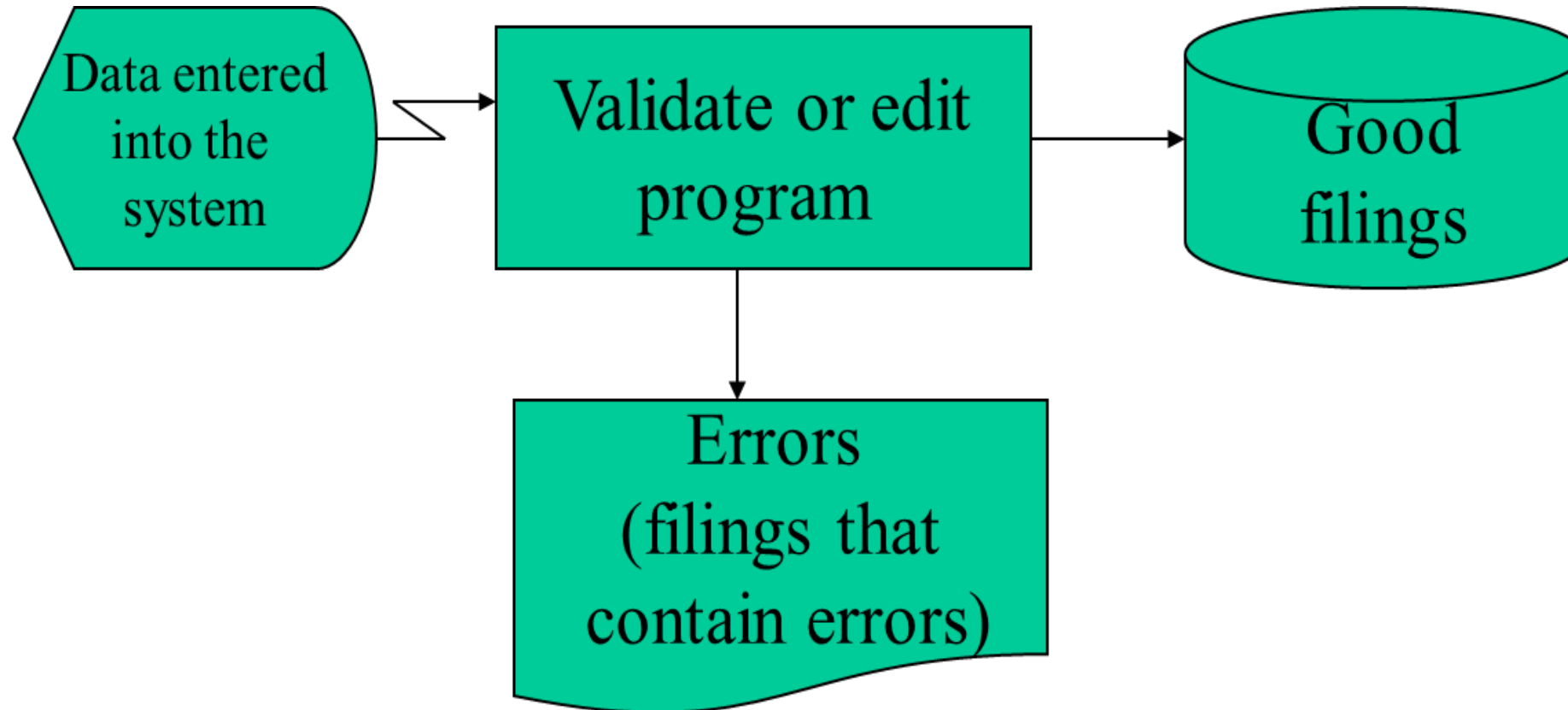
In this systems flowchart, I am showing data being entered from a screen. The validate/edit program is checking the data. Data that passes the tests will be written on the good transaction file. Data that contains errors are written to the screen. Other methodologies will be shown on the next few slides.

http://192.168.2.6:8080/IpasWeb