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

Creation of Another System

Perception of Bad Data Quality

Unwillingness to use the system
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?

Data Verification

We all make mistakes and so it’s important to check!

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

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
Best Practices – IPAS Data Capture

1. New App. Reception:
   - Paper-form to Electronic
   - Receive & Capture New Application
   - Validation & Verification

2. Amendments:
   - Paper-form to Electronic
   - Receive, Process Doc & Update Application
   - Validation & Verification

3. Assignments:
   - Paper-form to Electronic
   - Receive, Process Doc & Update Application
   - Validation & Verification

   IPAS Database

Create IP Record
## Data Standardization

<table>
<thead>
<tr>
<th>Unstandardized Data</th>
<th>Standardized Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Corp.</td>
<td>Microsoft Corporation</td>
</tr>
<tr>
<td>DataFlux Inc</td>
<td>DataFlux Incorporated</td>
</tr>
<tr>
<td>Windmill Co.</td>
<td>Windmill Company</td>
</tr>
<tr>
<td>MR RAMSAY SHONGE</td>
<td>Mr. Ramsay Shonge</td>
</tr>
<tr>
<td>Mister Ramsay Shonge</td>
<td>Mr. Ramsay Shonge</td>
</tr>
<tr>
<td>Mrster Robert Chiweza</td>
<td>Mr. Robert Chiweza</td>
</tr>
<tr>
<td>#(877)8463589 EXT250</td>
<td>877-846-3589 ext 250</td>
</tr>
<tr>
<td>8778463589 x250</td>
<td>877-846-3589 ext 250</td>
</tr>
<tr>
<td>Adams and Adams</td>
<td>Adams &amp; Adams</td>
</tr>
<tr>
<td>COCA-COLA</td>
<td>Coca Cola</td>
</tr>
<tr>
<td>Int’l Paper Company</td>
<td>International Paper Company</td>
</tr>
</tbody>
</table>
Defining Issues with your Data - Sample

Data Sources

- IPAS EDMS
- IPAS Database
- WIPO Databases
- Physical Files
- Other

Define Issues

- 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

- 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
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