Notes on parsing the PDF containing the FinCen spec for a rule that at least reduces the effort of transcribing the metadata, if not entirely automating the process

|  |  |  |
| --- | --- | --- |
| 1-2 | Record Type | 2 |
| 3-7 | Transaction Location Code | 5 |
| 8 | Multiple Transaction Locations (Item 22) | 1 |
| 9 | Type of Business Location (Item 23) | 1 |
| 10-64 | Transaction Location Legal Name (Item 24) | 55 |
| 65-119 | Doing Business As Name (Item 25) | 55 |
| 120-169 | Transaction Location Permanent Address  (Item 26) | 50 |
| 170-194 | Transaction Location City (Item 27) | 25 |
| 195-196 | Transaction Location State (Item 28) | 2 |
| 197-205 | Transaction Location ZIP Code (Item 29) | 9 |
| 206-207 | Transaction Location Country (Item 32) | 2 |
| 208-216 | Transaction Location EIN or SSN/ITIN (individual) (Item 30) | 9 |
| 217-226 | Transaction Location Business Phone Number (Item 31) | 10 |

**Rules**

**Import the table into Word**

Delete header row

Delete Description column 4 (base 1)

If column 1 contains a [-] symbol, delete it, and subsequent characters

Given the aim is to produce the format below…

For each row

prepend [{], print the integer in column 1, print [,]

print the integer in column 3, print [,]

print [“], print the description in Column 2, print [“},] (if it is the final row, do nnot print the trailing [,]

Convert the table to text (possibly before the “for each row” actions

|  |
| --- |
| {1,2,"Record Type"}, |
| {3,5,"Transaction sequence Number"}, |
| {8,1,"Type of Filing"}, |
| {9,14,"Document Control Number or BSA Identifier"}, |
| {23,8,"Date of Transaction"}, |
| {31,5,"Transaction Type"}, |
| {36,15,"Total Cash-in (Item 25)"}, |
| {51,15,"Cash-in: Deposit(s) (Item 25a)"}, |
| {66,15,"Cash-in: Payment(s)(Item 25b)"}, |
| {251,15,"Total Cash-out"} |