

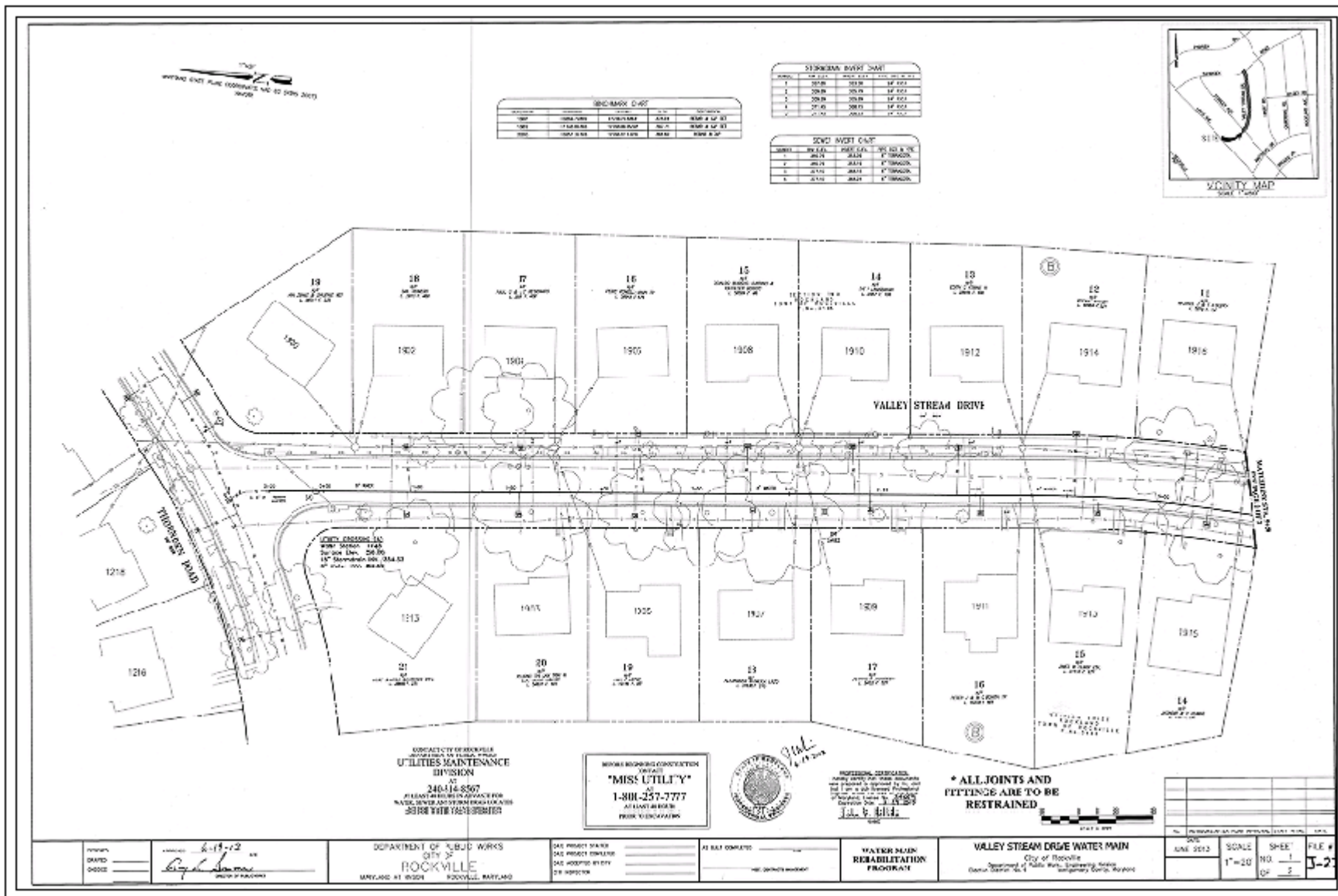
MrSID to PDF: Converting 13,000 files with ArcPy / Python



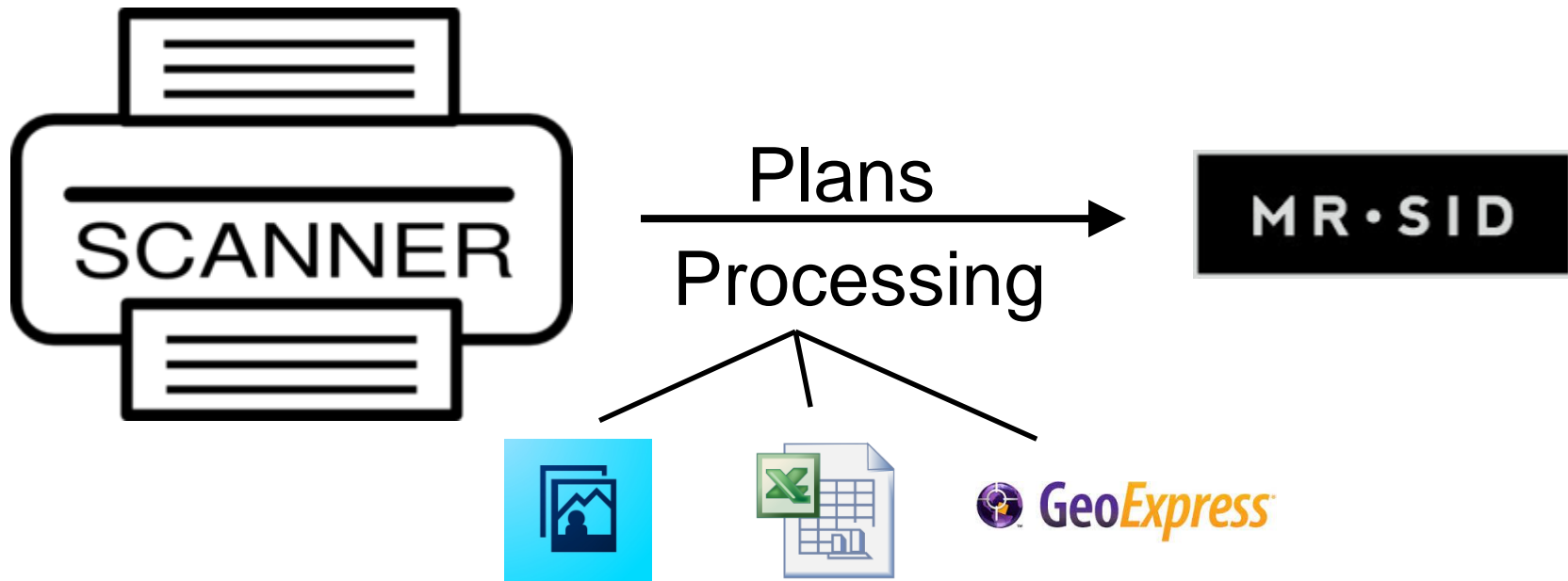
City of Rockville
Mike Onzay, Celine Yang, Dung Tran

Goals of Project

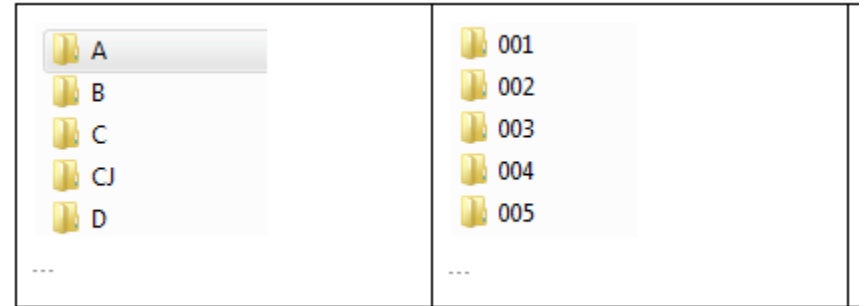
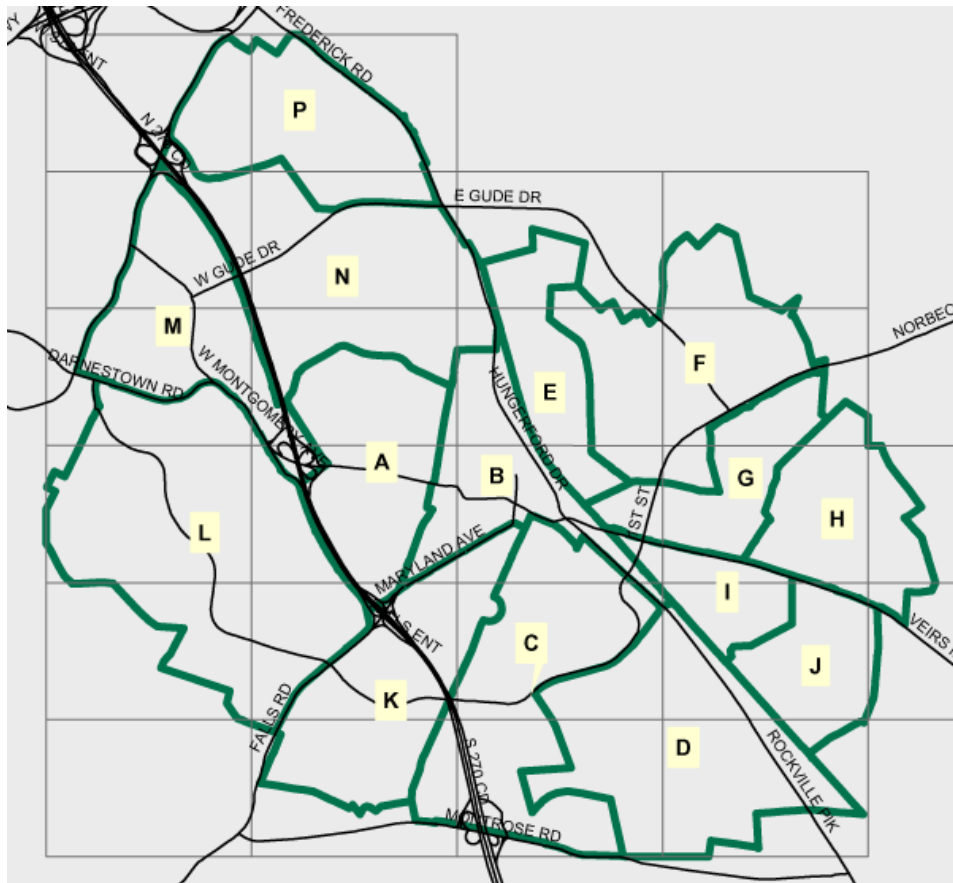
- Convert MrSID files to PDF
- PDF can be printed at correct drawing scale
- Retain coded filenames
- Create Engineering drawing sets
- Link to PDF from ArcMap
- Learn ArcPy and Python

















Internal Workflow



B-374-12-INSW-2014





-  B-374-01-OV-2014.sid
-  B-374-02-NO-2014.sid
-  B-374-03-O-2014.sid
-  B-374-04-O-2014.sid
-  B-374-05-O-2014.sid
-  B-374-06-O-2014.sid
-  B-374-07-NO-2014.sid
-  B-374-08-LO-2014.sid
-  B-374-09-LNO-2014.sid
-  B-374-10-NO-2014.sid
-  B-374-11-SW-2014.sid
-  B-374-12-INSW-2014.sid
-  B-374-13-NSW-2014.sid
-  B-374-14-NSW-2014.sid

MrSID + AutoCAD = PDF






- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z






- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

 B-002-01-DGIP-1950.sid

 B-002-01-DGIP-1950.pdf

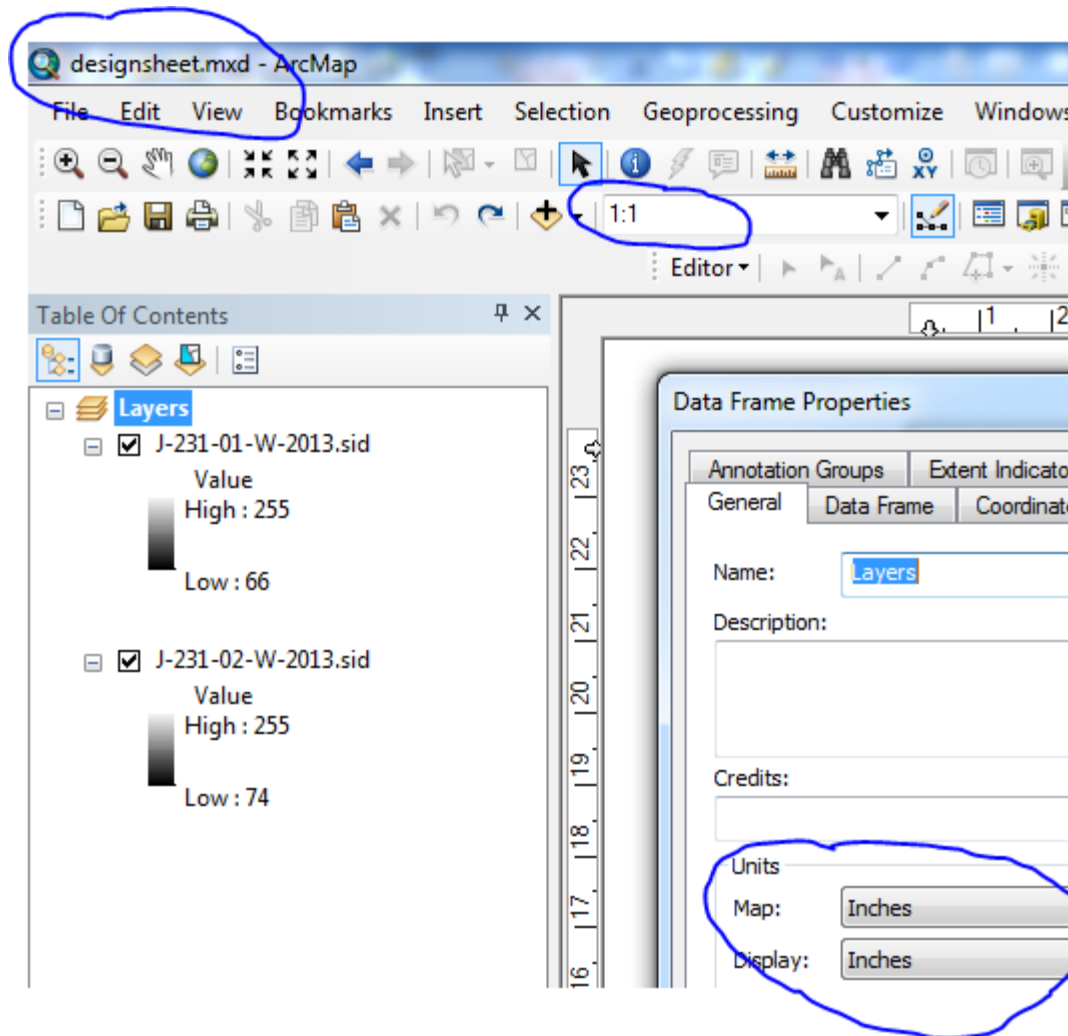
- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

-  B-374-01-OV-2014.sid
-  B-374-02-NO-2014.sid
-  B-374-03-O-2014.sid
-  B-374-04-O-2014.sid
-  B-374-05-O-2014.sid

-  B-374-01-OV-2014.pdf
-  B-374-02-NO-2014.pdf
-  B-374-03-O-2014.pdf
-  B-374-04-O-2014.pdf
-  B-374-05-O-2014.pdf

- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z

MrSID to PDF: Converting 13,000 files with ArcPy / Python



MrSID to PDF with Python inside ArcMap

```
import arcview  
import arcpy,os
```

```
mxd = arcpy.mapping.MapDocument("CURRENT")  
df = arcpy.mapping.ListDataFrames(mxd)[0]
```

```
for lyr in arcpy.mapping.ListLayers(mxd,"",df):  
    lyr.visible = "True"  
    lname = lyr.name  
    outFile = os.path.join(r"H:\DesignSheet\PDFs\\", lname.replace('.sid') + ".pdf")  
    arcpy.mapping.ExportToPDF(mxd, outFile, "PAGE_LAYOUT", 640, 480, 300,  
"BETTER", "RGB", "TRUE", "ADAPTIVE", "RASTERIZE_BITMAP", "", "",  
"LAYERS_ONLY", "FALSE")  
    lyr.visible = "False"  
del mxd
```


MrSID to PDF with Python outside ArcMap

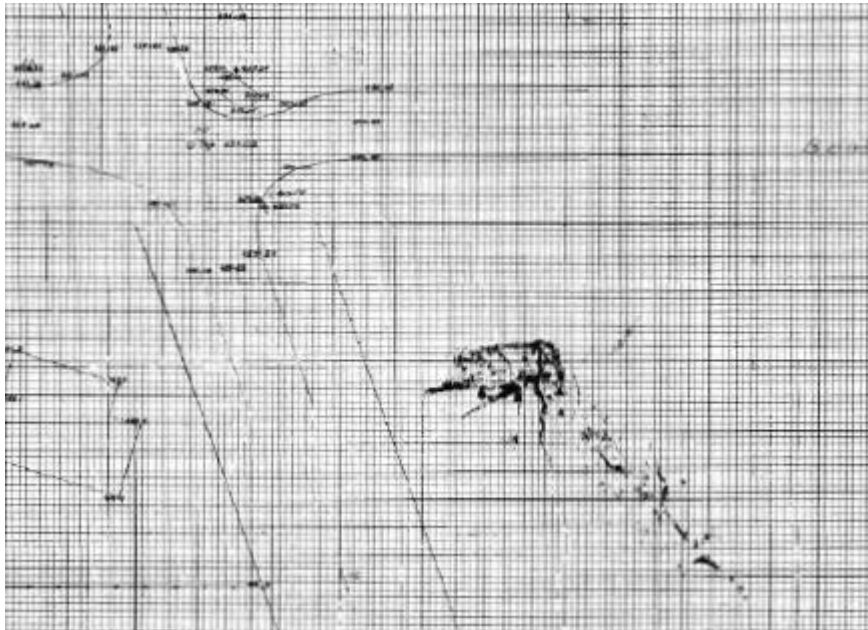
```
import arcview
import arcpy,os

mxd = arcpy.mapping.MapDocument(r"C:\GIS\designsheet.mxd")
df = arcpy.mapping.ListDataFrames(mxd)[0]
print df.scale
lyr = arcpy.mapping.Layer(r"\\gisserver2\DesignSheet\A\001\A-001-01-M-1956.sid")
arcpy.mapping.AddLayer(df,lyr)
print df.scale
lyr.visible = "True"
lname = lyr.name
outFile = r"C:\GIS\output\\" + lname.replace('.sid', '.pdf')
arcpy.mapping.ExportToPDF(mxd, outFile, "PAGE_LAYOUT", 640, 480, 300, "BETTER", "RGB",
"TRUE", "ADAPTIVE", "RASTERIZE_BITMAP", "", "", "LAYERS_ONLY", "FALSE")
lyr.visible = "False"
print outFile

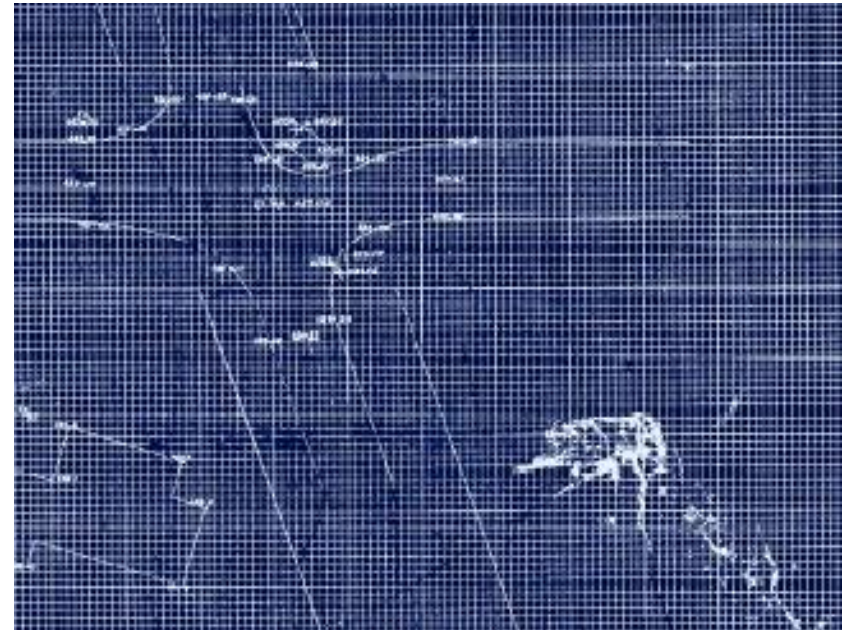
del mxd
```

MrSID to PDF with Python outside ArcMap

Expected Output



Unexpected Output



MrSID to PDF with Python outside ArcMap

Bug NIM-095188

Bug Number	NIM095188
Submitted	Sep 24, 2013 3:06 PM
Severity	Medium
Applies To	ArcGIS for Desktop
Version Found	10.1
Prog Language	N/A
Server Platform	Windows 7
Client Platform	Windows 7
Database	N/A
Locale	N/A
Status	Open: Submitted
Version Fixed	N/A
SP Fixed	N/A

Synopsis

Typing 'arcpy.mapping.AddLayer' in the Python window adds an unwanted, inverted blue colored raster to the map.

Additional Status Information

N/A

Alternate Solution

N/A

MrSID to PDF with Python outside ArcMap

```
arcpy.MakeRasterLayer_management(sourcefile, fnametemp)
```

```
arcpy.SaveToLayerFile_management(fnametemp, fnamesave)
```

```
addlayer = arcpy.mapping.Layer(fnamesave)
```

```
arcpy.mapping.AddLayer(df, addlayer, "Auto_ARRANGE")
```

```
arcpy.ApplySymbologyFromLayer_management (fnamesave, fnametemp)
```

MrSID to PDF with Python outside ArcMap



Bug NIM-099119

Bug Number	NIM099119
Submitted	Feb 19, 2014 8:04 PM
Severity	Medium
Applies To	ArcGIS for Desktop
Version Found	10.2.1
Prog Language	N/A
Server Platform	Windows 7
Client Platform	Windows 7
Database	N/A
Locale	N/A
Status	Open: Submitted
Version Fixed	N/A
SP Fixed	N/A

Synopsis

Unable to set the data frame scale to 1:1 in Python.

Additional Status Information

N/A

Alternate Solution

N/A

```
if df.scale != 1.0:
    df.scale = 0.999999
```

MrSID

+


ArcMap / ArcPy






=


PDF




- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z

- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

 B-002-01-DGIP-1950.sid

-  B-374-01-OV-2014.sid
-  B-374-02-NO-2014.sid
-  B-374-03-O-2014.sid
-  B-374-04-O-2014.sid
-  B-374-05-O-2014.sid

 B-002-01-DGIP-1950.pdf

-  B-374-01-OV-2014.pdf
-  B-374-02-NO-2014.pdf
-  B-374-03-O-2014.pdf
-  B-374-04-O-2014.pdf
-  B-374-05-O-2014.pdf

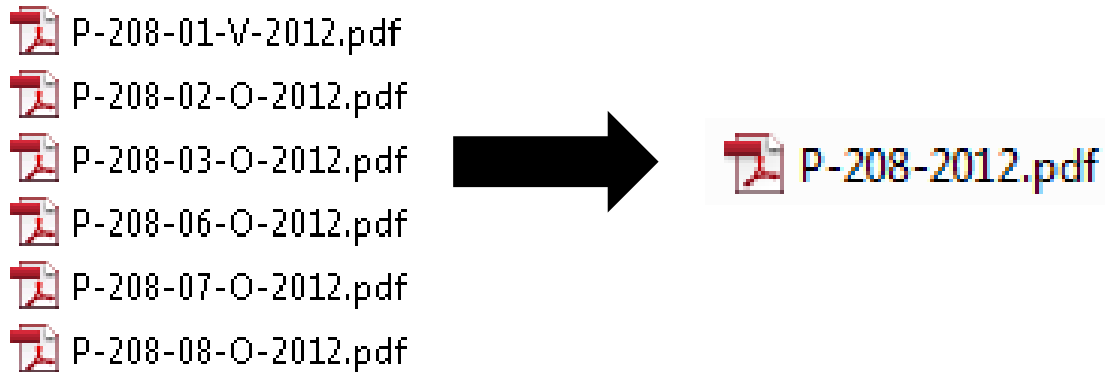
- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z

Goals of Project

- ✓ Convert MrSID files to PDF
- ✓ Retain coded filenames
- ✓ PDF can be printed at correct drawing scale
- Create Engineering drawing sets
- Link to PDF from ArcMap
- Learn ArcPy and Python

New Filename Syntax



PyPDF2

PyPDF2 is a pure-python PDF library capable of splitting, merging together, cropping, and transforming the pages of PDF files. It can also add custom data, viewing options, and passwords to PDF files. It can retrieve text and metadata from PDFs as well as merge entire files together.

Homepage

<http://mstamy2.github.io/PyPDF2/>

Combining PDFs – PyPDF2

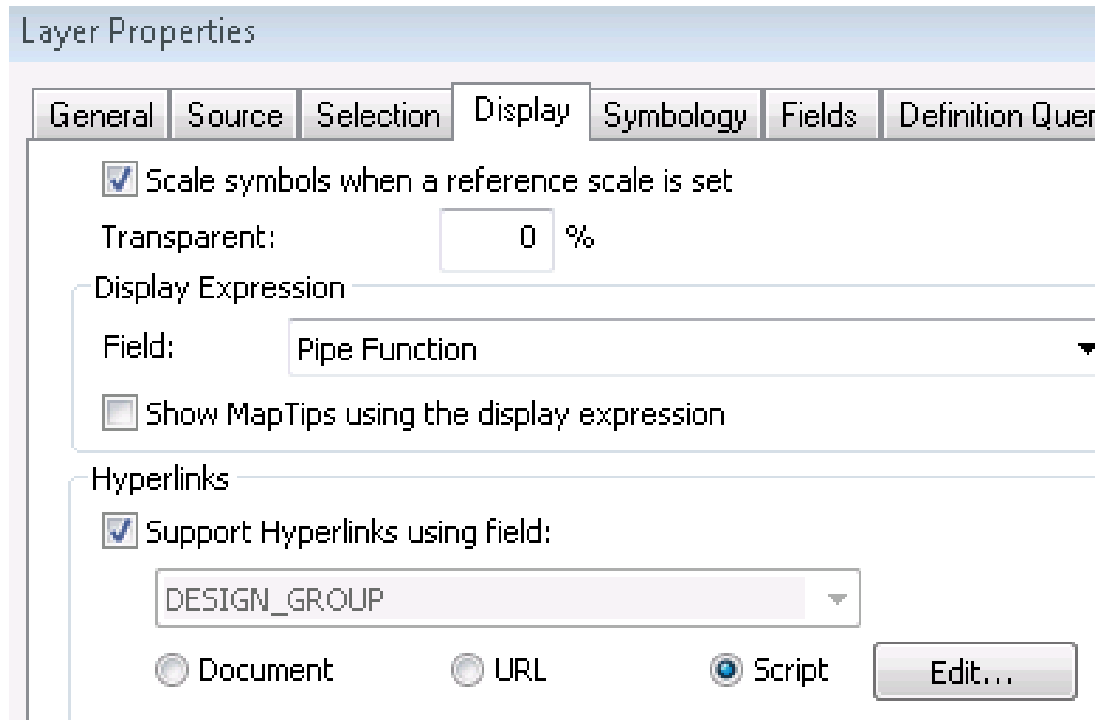
```
page = 0
merger = PdfFileMerger()
...
merger.append(open(os.path.join(PDF_source,pdf), 'rb'))
    merger.addBookmark(pdf, page)
    merger.setPageMode("/UseOutlines")
    page += 1
...
merger.addMetadata({u'/Author':u'City of Rockville'})
merger.addMetadata({u'/Keywords': u'public works, engineering'})
...
merger.write(open(os.path.join(PDF_letter,number), 'wb'))
```


Open PDF from ArcMap

The screenshot displays the ArcMap interface. On the left, the 'Table Of Contents' window shows a layer named 'Pipes (by function)' which is expanded to show 'Pipe Function' with three sub-categories: 'Distribution' (blue line), 'Fire' (red line), and 'Transmission' (magenta line). The central map area shows a network of pipes, with a magenta line (Transmission) highlighted. On the right, the 'Identify' window is open, showing the 'Pipes (by function)' layer selected. Below the layer name, two pipe IDs are listed: 'WP21930' and 'WP21925'. The 'Location' field shows coordinates: '1,277,581.619 514,503.329 Inches'. Below this is a table with the following data:

Field	Value
DIAMETER	8
DESIGN_GROUP	H195
YEAR_INSTALLED	2013

Open PDF from ArcMap



Open PDF from ArcMap

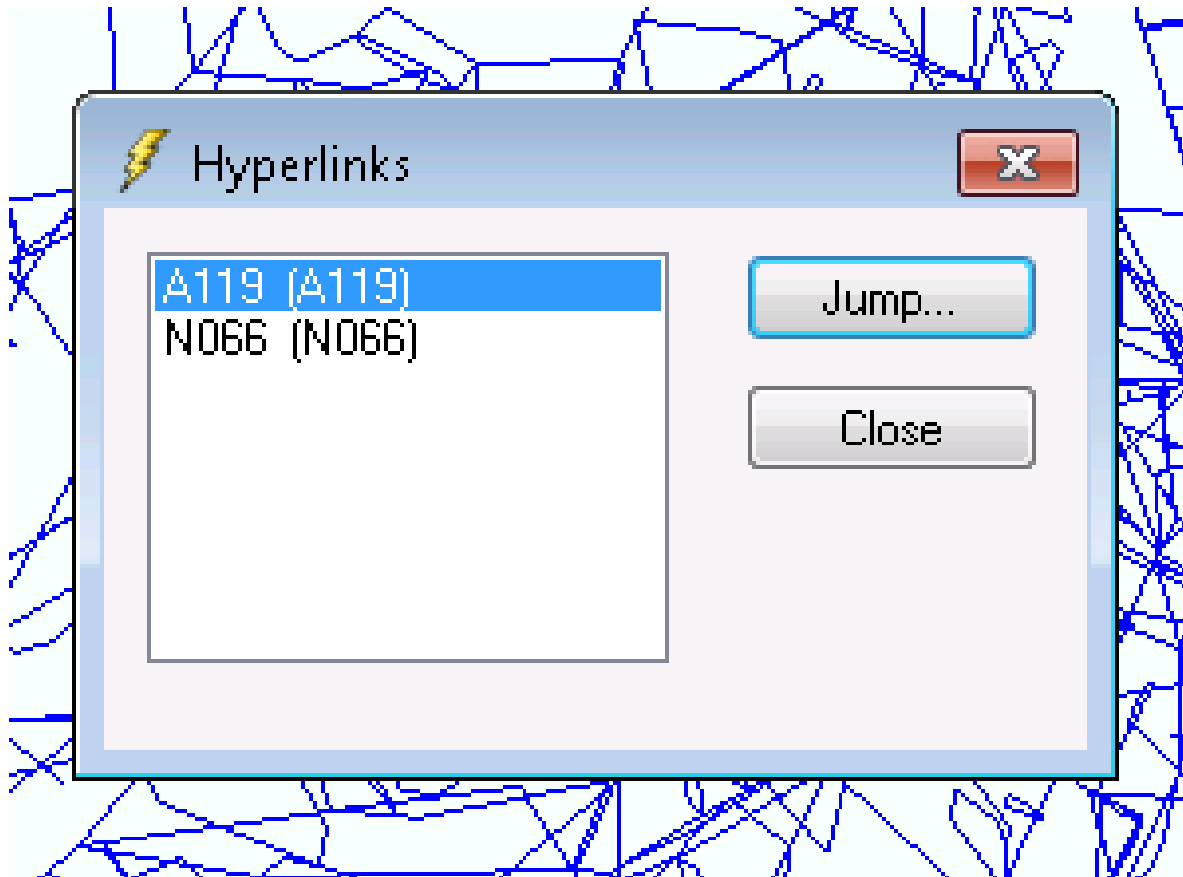
```
import webbrowser #only need this line inside ArcMap
import os
def OpenLink( [DESIGN_GROUP] ):
    path = r'\\files\Public Works\Design Sheets-PDF Converted'
    design = [DESIGN_GROUP]

    code_list = design.split(',')

    for code in code_list:

        if len(code) == 4:
            letter = code[0]
            number = code[1:4]
            for root, folder, files in os.walk(path):
                for pdf in files:
                    if pdf[:5] == '{0}-{1}'.format(letter, number):
                        os.startfile(os.path.join(path, letter, number, pdf))
```

Open PDF from ArcMap




Phase One Complete






- ✓ About 9,000 MrSID files were converted to PDF
- ✓ About 13,000 PDF files were combined into 4800 PDF drawing sets.
- ✓ Critical design sheet codes are stored as bookmarks in each PDF
- ✓ Can link to PDFs from ArcMap
- ✓ Can easily print one or more pages from PDF


MrSID + ArcMap
ArcPy / Python = PDF


- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z

- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

 B-002-01-DGIP-1950.sid

-  B-374-01-OV-2014.sid
-  B-374-02-NO-2014.sid
-  B-374-03-O-2014.sid
-  B-374-04-O-2014.sid
-  B-374-05-O-2014.sid

 B-002-1950.pdf

 B-374-2014.pdf

-  B-374-01-OV-2014.pdf
-  B-374-02-NO-2014.pdf
-  B-374-03-O-2014.pdf
-  B-374-04-O-2014.pdf
-  B-374-05-O-2014.pdf

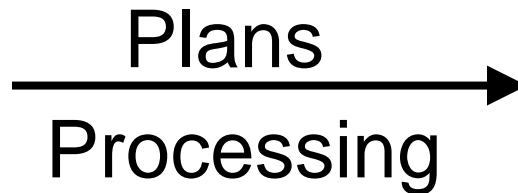
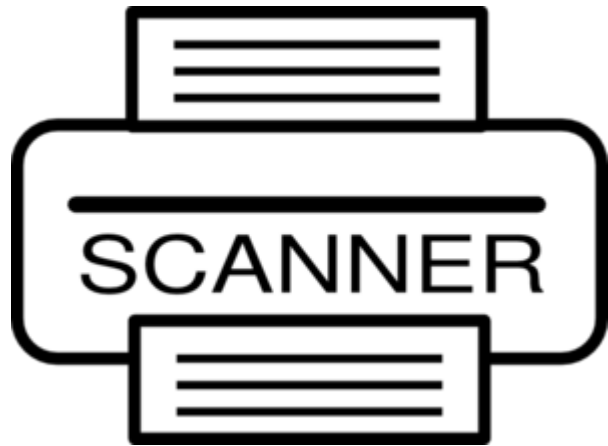
- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z

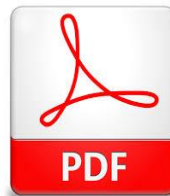
Phase Two – Moving Forward

- Need to adjust workflow for new files
- Need to accommodate revisions and as-builts into existing files
- Still need to convert ‘old/older/oldest’ files

Phase Two – Moving Forward



Still kept for georeferencing

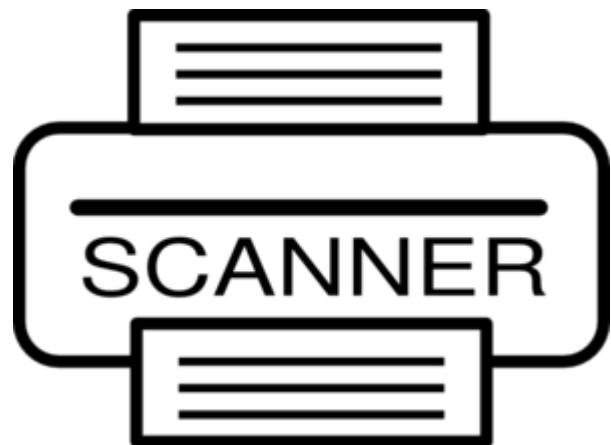


Single page PDFs

Manually moved to final folders

Burdensome dealing with as-builts and revisions

Phase Two – Moving Forward

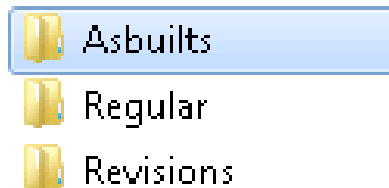


Plans
Processing



C-242-01-AV-2013.pdf

P-208-13-NO-2012_Rev2.pdf



Single page PDFs

Manually moved to final folders

~~Burdensome dealing with asbuilts and revisions~~ **Use Python!**

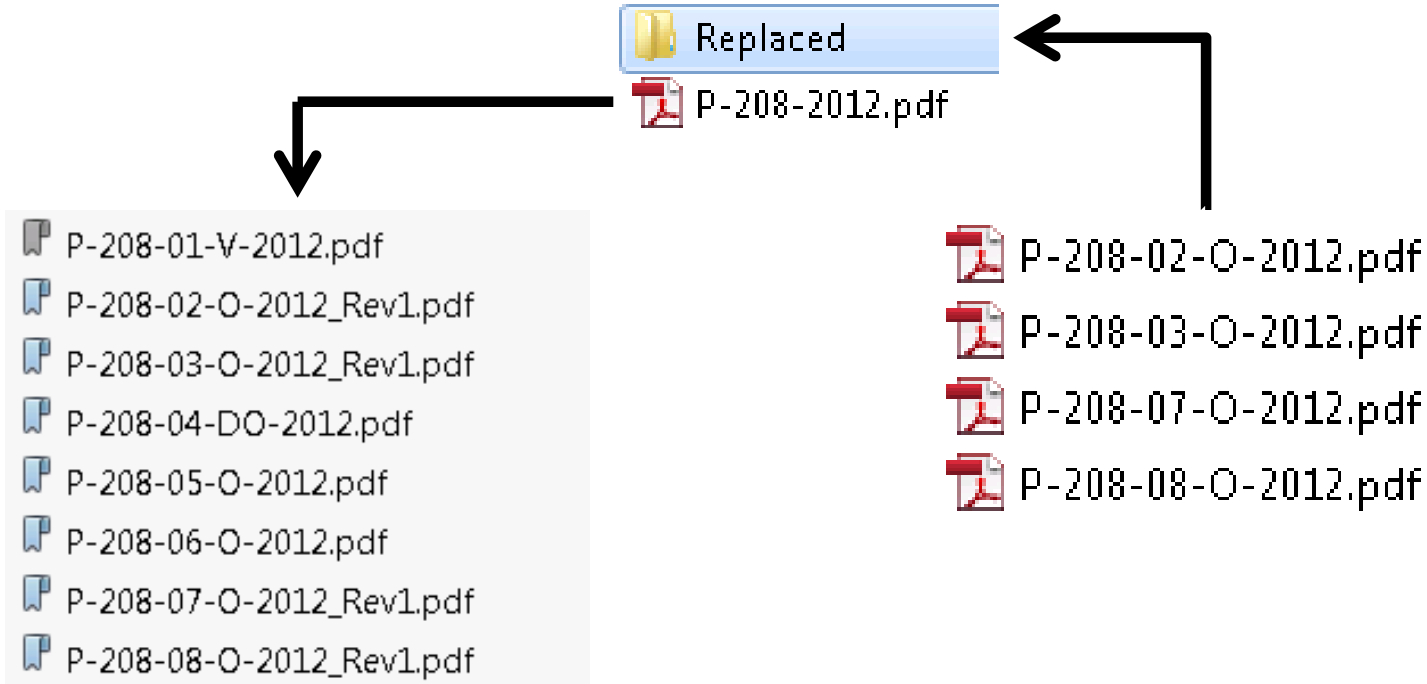
Phase Two – PyPDF2

Import os, shutil, datetime

From PyPDF2 import PdfFileReader, PdfFileWriter, PdfFileMerger

1. Check if existing PDF needs to be updated
2. Get bookmark names
3. Split single PDF into Multiple PDFs (pages)
4. Determine which pages need revisions
5. Replace pages
6. Move the old page into replaced folder
7. Merge updated pages + non-updated pages into single pdf
8. Write log files
9. Copy and overwrite updated PDFs to folders
10. Delete all staging files

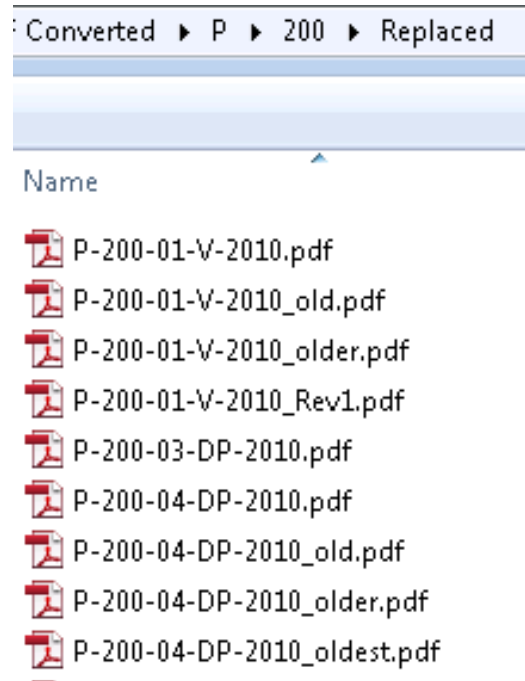
PyPDF2 – Revisions and Asbuilts



Phase Two – Complete

- Adjusted workflow for new files, revisions and as-builts
- Converted 1,117 ‘old/older/oldest’ files
- Matching MrSID and PDF folder/files

Phase Two – Complete



MrSID + ArcMap
ArcPy / Python = PDF

- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z

- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

- B-002-01-DGIP-1950.sdw
- B-002-01-DGIP-1950.sid
- B-374-01-OV-2014.sid
- B-374-02-NO-2014.sid
- B-374-03-O-2014.sid
- B-374-04-O-2014.sid
- B-374-05-O-2014.sid

- B-002-1950.pdf
- B-374-2014.pdf

- 001
- 002
- 003
- 004
- 005
- 006
- 007
- 008
- 009
- 010
- 011
- 374

- A
- B
- C
- CJ
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- Misc
- N
- O
- P
- RC
- Tanks
- Viewer
- WB
- WTP
- Z

Replaced

Replaced
P-208-2012.pdf

Lessons Learned

Python libraries whether proprietary (ArcPy) or open source (PyPDF2) make coding easier

Be ready to deal with exceptions.

Hidden files, such as Thumbs.db, are tricky to deal with when you don't know they are there.

Python sets are very cool

We learned how to pass a variable outside of a function

Use `str.replace(old, new[, count])` instead of `str.rstrip([chars])`

You cannot use Acrobat Pro to replace or add a page on a PDF that was touched by PyPDF2 or the internal structure of the bookmarks will break.

If you are stuck ask for help -- offline. Brainstorming with a real person can be fun.

Document your code early and often

MrSID to PDF: Converting 13,000 files with ArcPy / Python

Questions ?

City of Rockville
Mike Onzay, Celine Yang, Dung Tran