Steven D. Straka
6/02/2005
The Sanborn Map Company, Inc.

Sanborn's Metadata Creation Process for USDA NAIP Projects

In 2004 Sanborn won contract with the U.S. Department of Agriculture (USDA) for the National Agriculture Imagery Program (NAIP) which consisted of four states. Each state had a delivery of DOQQ's and County Compressed Mosaics (CCMs) for each county. Once the imagery was captured the turn around for a final delivery was very short and required that most processes be automated. Each CCM had to be accompanied by a metadata file and a shapefile showing the coverage and DOQQ's make-up of that CCM as well as a metadata file for that shapefile. This paper addresses how Sanborn used ArcObjects 9.0, ArcCatalog 9.0, VB 6.0 and the latest version of Perl to create and maintain this metadata. I will address how Sanborn used the USGS metadata parser (mp) to verify the metadata before delivery to the USDA. How ArcMap 9.0 was used to visually verify the MrSID CCM files before shipping will also be discussed.

BACKGROUND

In 2004 the USDA awarded Sanborn four states under its National Agriculture Imagery Program (NAIP). The four states included Arizona, Colorado, South Dakota, and North Carolina. These were a combination of 1 meter imagery states and 2 meter imagery states. The major deliveries for a NAIP state are the DOQQ center TIFF files that make up the state, a MrSID Compressed County Mosaic (CCM) for each county in the state, and the metadata the goes along with each CCM .

NAIP annually acquires current and accurate imagery of all agricultural lands of the continental United States and delivers the data to USDA Service Centers. The imagery is used for several purposes, including crop management and determining farm and tract boundaries. It's imperative that the imagery acquisition process be timely and accurate to met these requirements

This paper will address the creation of that metadata. I will cover the composition of the metadata, the tools and processes used to create that metadata. I will also cover the QC tools and processes used in verifying the correctness of metadata. Issues that were encountered in creation of metadata as well as the QC of the metadata using the USGS metadata parser (mp) will also be touch upon.

Sanborn used a database called LiftTrack to collect and archive much of the data needed for the metadata creation. I will talk about the use of LiftTrack and how it was

used in the production of metadata. This database is an enterprise Microsoft SQL server (MS-SQL) 2000 database running in a windows environment.

The last items that I will discuss are the enhancements that need to be made in order to make this process even more streamlined for Sanborn. These will include near term and future wish list items.

DESIGN CONSIDERATION

The metadata deliveries for a NAIP state were complex and had many rules as to how the metadata was created including how the format was to be structured, but primarily what was to be included and not included in the metadata. Sanborn wanted the metadata creation to be robust yet simple. The main goal was the correctness of the metadata. It was important that a delivery not be rejected due to "bad" metadata. Using ArcCatalog metadata editor alone was out of the question as there were just too many fields to be populated by hand which would increase the chance manual errors in the metadata creation.

Sanborn needed other items such as QC scripts, MrSID input scripts, etc. in addition to metadata to be produced at the same time that made logical sense to add to the metadata translators. The first thing needed was the MrSID script and input text file that would be used as input to the MrSID compressor. Other items required were scripts and input text to drive the metadata creation process.

DEVELOPMENT / CODE EXAMPLES

Sanborn decided that VB with ArcObjects would be the method used for automating the metadata creation process. VB offered the development of quick GUIs and a well documented interface into ArcObjects. The need for Perl arrived after the ArcObjects development was done and there still remained a few strings that needed replaced. Perl with its strong string handling abilities seemed to make it the best choice. The USGS MP program was just a command line program that Sanborn was able to script and run within a bash shell. Cygwin Bash was chosen to allow all the processes to be scripted and hide the Perl and mp programs from the user. The Shapefile metadata and the CCM metadata creation steps and examples are included below.

Shapefile, Shapefile metadata and MrSID script creation

First, a GUI was developed that allowed the user to select the state and county that they were going to develop an MrSID CCM for:

Figure 1 is the main GUI for creating the Shapefile, Shapefile metadata, MrSID script and the QC scripts. Notice that the user must select the UTM zone that the CCM will be delivered in as well as the state and county.

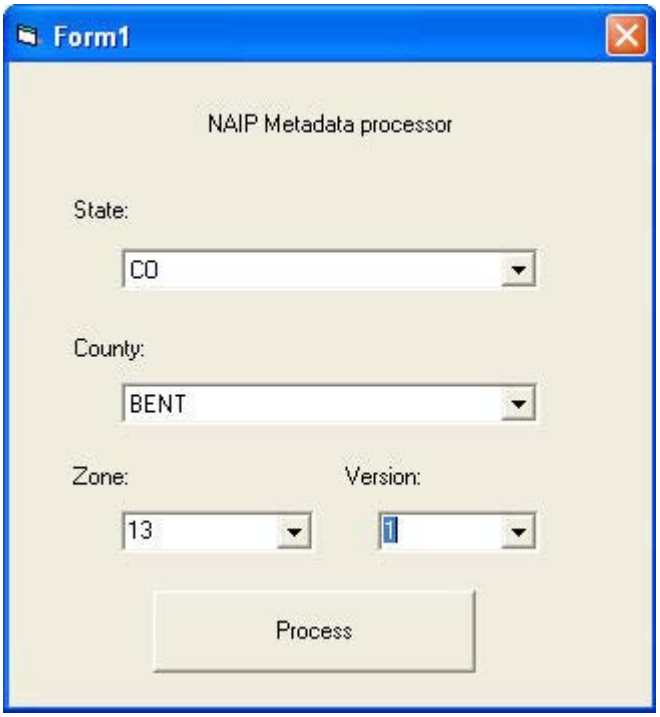

Figure 1: Shapefile and MrSID script GUI

Figure 2 shows part the text file that is created from out LiftTrack database and is used by the application to create the Shapefile, Shapefile metadata, MrSID script and the QC scripts. The key field that is used in linking this table to the DOQQ grid from which the shapefile is created is the DOQQ_ID.

```
"USDANAME","DOQQ","IDAT","QKEY","BCON","DOQQ_ID","STATE","CCM","QQTILENAME","QQNAME","UTM_ZONE","FIPS_COUNTY","IMAGE_ID", "INCLUDED"
"1026w-0438","n3710219.sw","2004-07-14","373730N1024115W","NC","37102-F6-SW","CO","BENT","n_3710219_sw_13_2_20040714.tif","HASSER RANCH SW","13","8011","070465455","Y"
"1026w-0440","n3710219.nw","2004-07-14","374115N1024115W","NC","37102-F6-NW","CO","BENT","n_3710219_nw_13_2_20040714.tif","HASSER RANCH NW","13","8011","070465457","Y"
"1026w-0442","n3710211.sw","2004-07-14","374500N1024115W","NC","37102-G6-SW","CO","BENT","n_3710211_sw_13_2_20040714.tif","CAT CREEK SW","13","8011","070465459","Y"
"1026w-0444","n3710211.nw","2004-07-14","374845N1024115W","NC","37102-G6-NW","CO","BENT","n_3710211_nw_13_2_20040714.tif","CAT CREEK NW","13","8011","070465461","Y"
"1026w-0446","n3710203.sw","2004-07-14","375230N1024115W","NC","37102-H6-SW","CO","BENT","n_3710203_sw_13_2_20040714.tif","CAT CREEK NW SW","13","8011","070465463","Y"
"1026w-0448","n3710203.nw","2004-07-14","375615N1024115W","NC","37102-H6-NW","CO","BENT","n_3710203_nw_13_2_20040714.tif","CAT CREEK NW NW","13","8011","070465465","Y"
"1026w-0450","n3810259.sw","2004-07-14","380000N1024115W","NC","38102-A6-SW","CO","BENT","n_3810259_sw_13_2_20040714.tif","LAMAR WEST SW","13","8011","070465467","Y"
"1026w-0452","n3810259.nw","2004-07-14","380345N1024115W","NC","38102-A6-NW","CO","BENT","n_3810259_nw_13_2_20040714.tif","LAMAR WEST NW","13","8011","070465469","Y"
"1026w-0454","n3810251.sw","2004-07-11","380730N1024115W","NC","38102-B6-SW","CO","BENT","n_3810251_sw_13_2_20040711.tif","WILEY SW","13","8011","070465471","Y"
"1026w-0456","n3810251.nw","2004-07-11","381115N1024115W","NC","38102-B6-NW","CO","BENT","n_3810251_nw_13_2_20040711.tif","WILEY NW","13","8011","070383874","Y"
```

Figure 2: County Text file from LiftTrack

Figure 3 is an example of part of the MrSID input text script that is created as part of the above process. This full directory path is given so that this process may be distributed to any of the Sanborn worker machines.

```
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B1-NW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-C1-SW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-C2-SE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B2-NE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B3-NW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-C3-SW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-C4-SE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B4-NE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B2-NW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-C2-SW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-C3-SE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B3-NE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-A1-NW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B1-SW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B2-SE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-A2-NE.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-A3-NW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B3-SW.tif
//Sanborn/Projects/NAIP/CO/13/BENT/deliv/38103-B4-SE.tif
```

Figure 3: MrSID input Script

Figure 4 shows part of the Shapefile and part of the table for that Shapefile created while running the above process. All of the attributes for the table came from the table produced from the LiftTrack database. Please note the date in the IDAT field, which is the date that the imagery was captured.

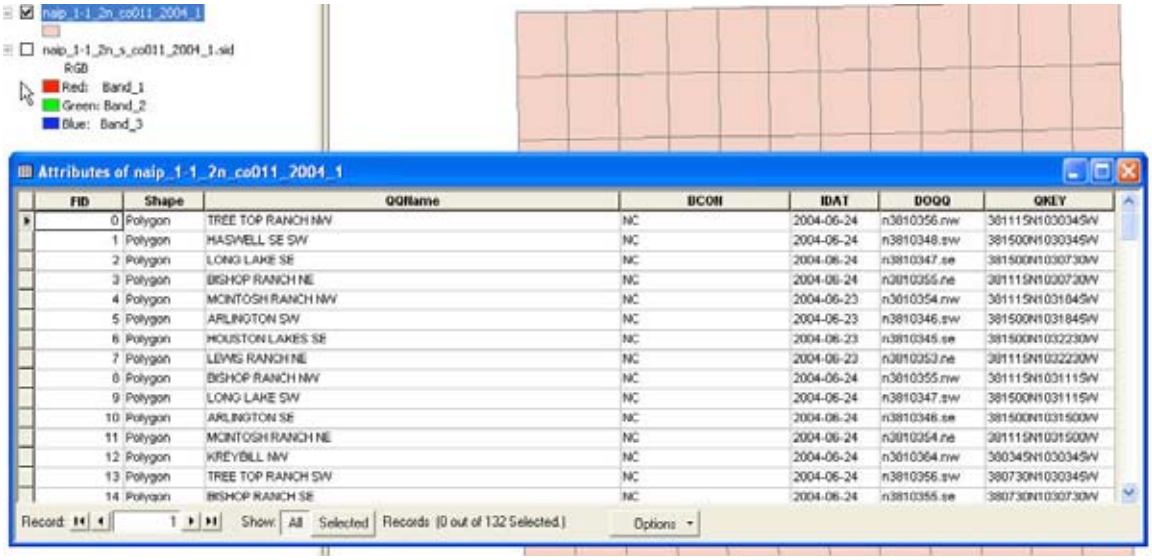

Figure 4: Shapefile and Table for Bent County

Figure 5 in an example of the Shapefile metadata that is created by the above process. You can see that both the shapefile table above and the metadata Point_and_Vector_Object_Count both show 132 polygons for Bent County.

```
      this shape file data set.
      Source_Used_Citation_Abbreviation: NAIP
      Process_Date: 20040710
Spatial_Data_Organization_Information:
  Indirect_Spatial_Reference: None
  Direct_Spatial_Reference_Method: Vector
  Point_and_Vector_Object_Information:
    SDTS_Terms_Description:
      SDTS_Point_and_Vector_Object_Type: G-polygon
      Point_and_Vector_Object_Count: 132
Spatial_Reference_Information:
  Horizontal_Coordinate_System_Definition:
    Planar:
      Grid_Coordinate_System:
        Grid_Coordinate_System_Name: Universal Transverse Mercator
        Universal_Transverse_Mercator:
          UTM_Zone_Number: 13
          Transverse_Mercator:
            Scale_Factor_at_Central_Meridian: 0.999600
            Longitude_of_Central_Meridian: -105.000000
            Latitude_of_Projection_Origin: 0.0
            False_Easting: 500000
            False_Northing: 0.0
      Planar_Coordinate_Information:
        Planar_Coordinate_Encoding_Method: coordinate pair
        Coordinate_Representation:
          Abscissa_Resolution: 0.000128
          Ordinate_Resolution: 0.000128
        Planar_Distance_Units: meters
    Geodetic_Model:
      Horizontal_Datum_Name: North American Datum of 1983
      Ellipsoid_Name: Geodetic Reference System 80
      Semi-major_Axis: 6378137.000000
      Denominator_of_Flattening_Ratio: 298.257222
Entity_and_Attribute_Information:
  Detailed_Description:
    Entity_Type:
      Entity_Type_Label: dBase file (.dbf)
      Entity_Type_Definition:
        The dBASE file that stores the attribute information of
        features. When a shapefile is added as a theme to a view,
        this file is displayed as a feature table.
      Entity_Type_Definition_Source: ESRI OnLine Help
    Attribute:
      Attribute_Label: QQName
      Attribute_Definition:
        Digital Orthophoto Quarter Quadrangle map name includes
        the quadrant locator e.g.(NE,NW,SE,SW).
      Attribute_Definition_Source: USDA-FSA Aerial Photography Field Office
      Attribute_Domain_Values:
        Codeset_Domain:
          Codeset_Name: Geographic Names Information System
          Codeset_Source: U.S. Geological Survey
    Attribute:
      Attribute Label: IDAT
```

Figure 5: Shapefile metadata for Bent county Colorado

CCM metadata and QC script creation

To make the CCM metadata creation very straight forward, a second GUI was developed which allowed the user to select the information needed to create the CCM and required AUX file that was associated with that CCM. Below is an example of that GUI:

Figure 6 is the GUI for the CCM creation. A spatial reference is supplied so that an AUX file can first be built for the MrSID file before applying the template. Notice also that the Zone and FIPS code must be provided as this is needed for the metadata.

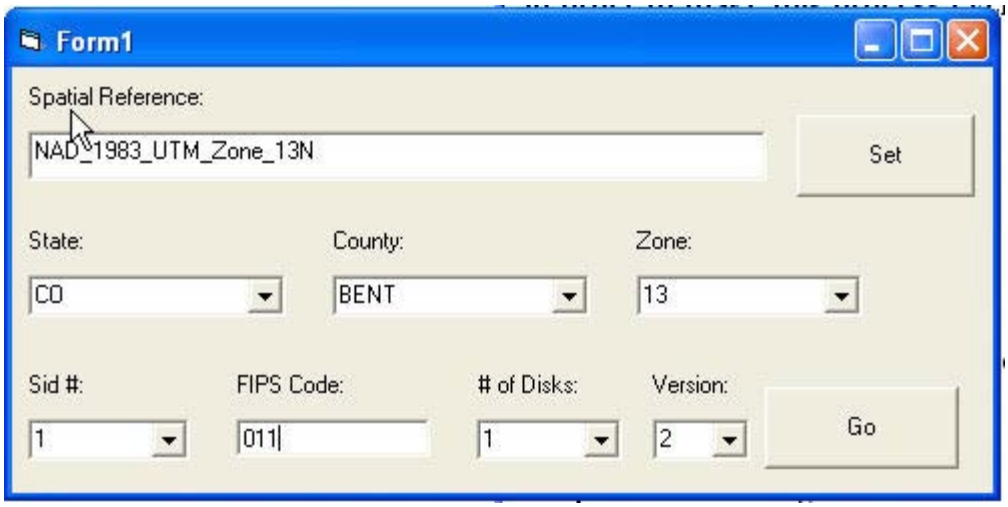

Figure 6: CCM GUI

Figure 7 is an example of part the of the DOQQ lineage input text file that is used as input to create the CCM metadata. This input file was created by the Shapefile process using the information that was provided by LiftTrack.

```
Source_Information:
  Source_Citation:
    Citation_Information:
      Originator: Aerial Photography Field Office
      Title: n3810356.nw
      Publication_Date: 2004-06-24
  Source_Scale_Denominator: 40000
  Type_of_Source_Media:          CD
  Source_Time_Period_of_Content:
    Time_Period_Information:
        Single_Date/Time:
          Calendar_Date: 2004
    Source_Currentness_Reference: Majority Aerial Photography Date
  Source_Citation_Abbreviation: MrSID compressed image
  Source_Contribution: 3.75' quarter quadrangle used in Mosaicked County Image
Source_Information:
  Source_Citation:
    Citation_Information:
      Originator: Aerial Photography Field Office
      Title: n3810348.sw
      Publication_Date: 2004-06-24
  Source_Scale_Denominator: 40000
  Type_of_Source_Media:          CD
  Source_Time_Period_of_Content:
    Time_Period_Information:
        Single_Date/Time:
          Calendar_Date: 2004
    Source_Currentness_Reference: Majority Aerial Photography Date
  Source_Citation_Abbreviation: MrSID compressed image
  Source_Contribution: 3.75' quarter quadrangle used in Mosaicked County Image
```

Figure 7: CCM lineage input text file for Bent county Colorado

Figure 8: A sample of a MrSID compress county mosaic (CCM) for Bent county Colorado.

Figure 8: MrSID CCM for Bent compressed 50:1

Figure 9 is the shapefile overlaying the CCM. This is one of many QC step to very completeness and correctness of the spatial reference of both datasets.

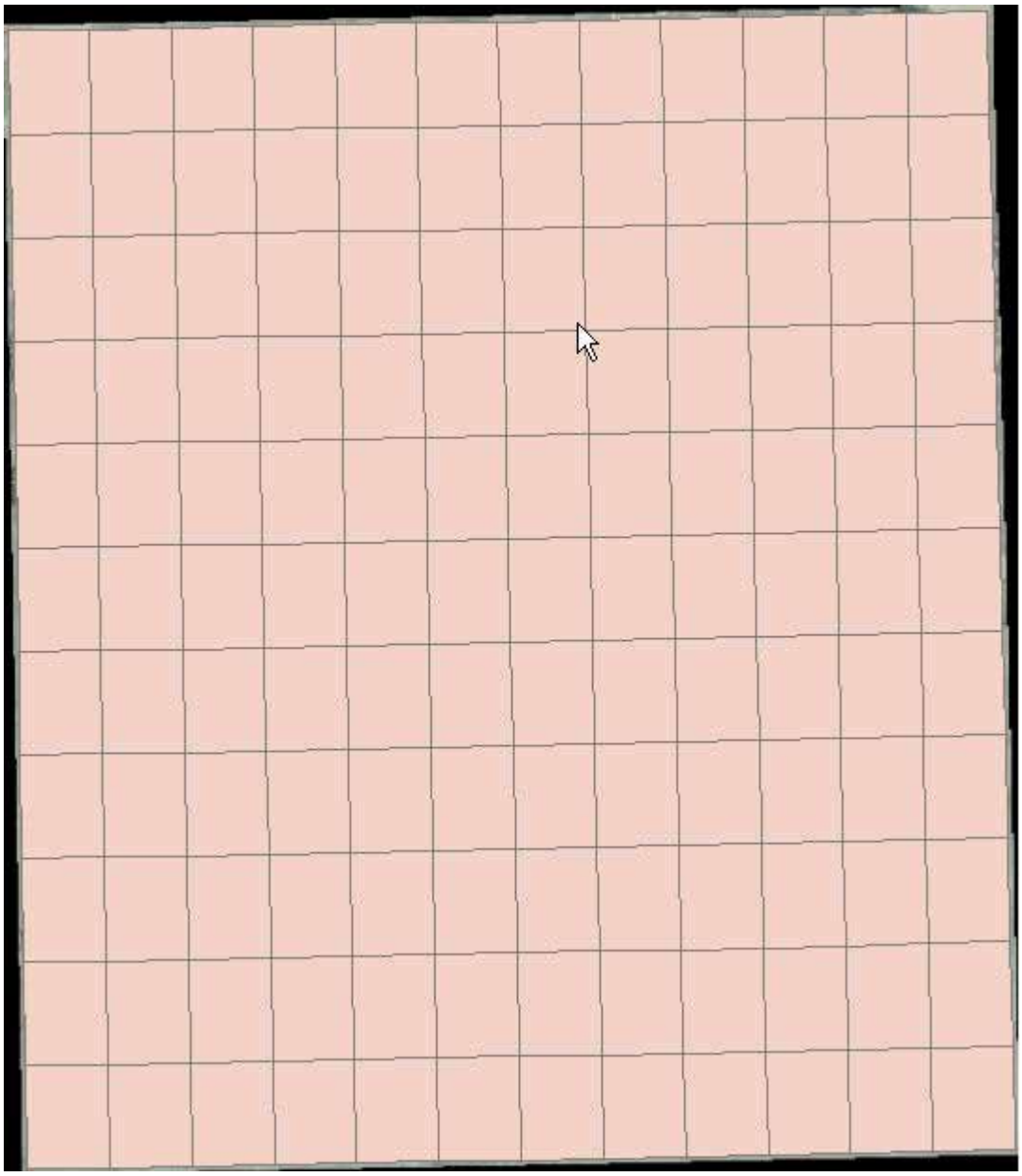

Figure 9: Shapefile for Bent, Notice the image behind the Shapefile

Figure 10 is an example of the part of CCM metadata for Bent county Colorado. Notice the FIPS code gets put in as part of the Place_Keyword. Also note the Lineage section which was replaced be the input text file shown above.

```
      Theme_Keyword: Aerial Compliance
      Theme_Keyword: Farming
    Place:
      Place_Keyword_Thesaurus: Geographic Names Information System
      Place_Keyword: CO
      Place_Keyword: BENT
      Place_Keyword: BENT CO
      Place_Keyword: 08011
  Access_Constraints: None
  Use_Constraints:
    Users should be aware that this is an interim release.  The
    interim compressed county mosaic may contain defects and have
    horizontal accuracy less than the specified tolerances.
    Defective imagery may be replaced within one year of image
    acquisition date.
  Point_of_Contact:
    Contact_Information:
      Contact_Organization_Primary:
        Contact_Organization: USDA-FSA Aerial Photography Field Office
      Contact_Address:
        Address_Type: mailing and physical address
        Address: 2222 West 2300 South
        City: Salt Lake City
        State_or_Province: Utah
        Postal_Code: 84119-2020
        Country: USA
      Contact_Voice_Telephone: 801-975-3500
      Contact_Facsimile_Telephone: 801-975-3529
  Browse_Graphic:
    Browse_Graphic_File_Name: None
    Browse_Graphic_File_Description: None
    Browse_Graphic_File_Type: None
  Native_Data_Set_Environment: Microsoft Windows 2000 Version 5.0 (Build 2195) Service Pack 4;
Data_Quality_Information:
  Logical_Consistency_Report:
    NAIP 3.75 minute tile file names are based
    on the USGS quadrangle naming convention.
  Completeness_Report: None
  Positional_Accuracy:
    Horizontal_Positional_Accuracy:
      Horizontal_Positional_Accuracy_Report: FSA Digital Orthophoto Specs.
    Vertical_Positional_Accuracy:
      Vertical_Positional_Accuracy_Report: N/A  2d only
  Lineage:
    Source_Information:
      Source_Citation:
        Citation_Information:
          Originator: Aerial Photography Field Office
          Title: n3810356.nw
          Publication_Date: 2004-06-24
      Source_Scale_Denominator: 40000
      Type_of_Source_Media:         CD
      Source_Time_Period_of_Content:
        Time_Period_Information:
            Single_Date/Time:
              Calendar_Date: 2004
        Source_Currentness_Reference: Majority Aerial Photography Date
      Source_Citation_Abbreviation: MrSID compressed image
      Source_Contribution: 3.75' quarter quadrangle used in Mosaicked County Image
    Source_Information:
      Source_Citation:
        Citation_Information:
          Originator: Aerial Photography Field Office
          Title: n3810348.sw
          Publication_Date: 2004-06-24
      Source_Scale_Denominator: 40000
      Type_of_Source_Media:         CD
      Source_Time_Period_of_Content:
        Time_Period_Information:
```

Figure 10: CCM metadata example

The code need to handle the generation of both the Shapefile and the CCM metadata was written using ArcCatalog Metadata code and a combination of Perl code with Bash scripts. Below are a couple of VB code snippets that I felt may be of some interest.

Applying Metadata Template

```
Dim m_pExportTXT As ExportMPTXT
Dim pExport As IMetadataExport
Dim m_pImportTXT As ImportMPTXT
Dim pImport As IMetadataImport
Dim pEnumDSName As IEnumDatasetName
Dim pWS As IWorkspace
Dim pDataset As IDataset


Set pFWS = pWkSpFact.OpenFromFile(WkSpLoc, 0)
Set MD_FC = pFWS.OpenFeatureClass(NAIPName & ".shp")

Set pDataset = MD_FC
Set pWS = pDataset.Workspace
Set pEnumDSName = pWS.DatasetNames(esriDTFeatureClass)
Set pDSName = pEnumDSName.Next

If pDSName.Name = NAIPName Then
Else
  Set pDSName = pEnumDSName.Next
End If

Set pMD = pDSName

'***Create or Synch metadata
pMD.Synchronize esriMSAAlways, 1

Set m_pExportTXT = New ExportMPTXT
Set pExport = m_pExportTXT

'***Export the metadata to a local directory - local tmplate copy for Bndy Coords.
pExport.Export pMD, WkSpLoc & "\" & NAIPName & ".tmplate"

Set m_pImportTXT = New ImportMPTXT
Set pImport = m_pImportTXT

'***Import a local xml document to metadata
 pImport.Import MetadataLoc & "\" & "shapefile_metadata_template.txt", pMD

Set m_pExportTXT = New ExportMPTXT
Set pExport = m_pExportTXT

'***Export the metadata to a local directory
pExport.Export pMD, WkSpLoc & "\" & NAIPName & ".met"
```

```
'*** Set the file for the Check script...
  LogfsoCHKTxt.CreateTextFile WkSpLoc & "\" & "CheckShapefile.s", True, False
  Set LogfleCHKTxt = LogfsoCHKTxt.GetFile(WkSpLoc & "\" & "CheckShapefile.s")
  LogfleCHKTxt.OpenAsTextStream(ForAppending).WriteLine ("../../mp.exe " & NAIPName & ".met")
```

Link and Join:

```
'***Create the MemoryRelationshipClass that defines what is to be joined
Dim pMemRelClassFact As IMemoryRelationshipClassFactory
Set pMemRelClassFact = New MemoryRelationshipClassFactory
Dim pRelClass As IRelationshipClass
Dim pTblFact As IWorkspaceFactory
Dim pTblWorkspace As IWorkspace

'*** Perform the link based on DOQQ_ID Number
Set pTblFact = New TextFileWorkspaceFactory
Set pTblWorkspace = pTblFact.OpenFromFile(CCMTextFileLoc & "\" & Form1.Combo1.Text, 0)
Set pFeatWS = pTblWorkspace
Set pTable = pFeatWS.OpenTable(ASISCCMName & ".txt")
RowCnt = pTable.RowCount(Nothing)
Set pRelClass = pMemRelClassFact.Open("CCM_Join", DOQQ_FC, _
 "DOQQ_ID", pTable, "DOQQ_ID", "forward", "backward", esriRelCardinalityOneToOne)

'*** Perform the join
Dim pRelQueryTableFact As IRelQueryTableFactory
Dim pRelQueryTab As ITable
Set pRelQueryTableFact = New RelQueryTableFactory
Set pRelQueryTab = pRelQueryTableFact.Open(pRelClass, True, Nothing, Nothing, "", True, True)

Dim pQueryFilter As IQueryFilter

Set pQueryFilter = New QueryFilter
pQueryFilter.WhereClause = "CCM = '" & ASISCCMName & "'" & "and INCLUDED = 'Y'"
Set pCursor = pRelQueryTab.Search(pQueryFilter, True)
```

Spatial Reference:

```
'***Open the raster dataset
Set pRasterDataset = pRasterWs.OpenRasterDataset(NAIPName & ".sid")
If pRasterDataset Is Nothing Then
   MsgBox "Error Opening MrSID file: " & NAIPName
   Exit Sub
End If

Dim pSchemaEdit As IGeoDatasetSchemaEdit
Set pSchemaEdit = pRasterDataset
pSchemaEdit.AlterSpatialReference m_SR
```

QC of the metadata file(s) was a very important step in the process and was required by the USDA. NAIP required that at least the USGS Metadata Parser program be run and pass on all metadata. MP generates a textual report indicating errors in the metadata, primarily in the structure but also in the values of some of the scalar elements (that is, those whose values are restricted by the standard). Below is an example of the script used to check a CCMs metadata before shipping. This script was created automatically as part of the first application that was run.

Check CCM Script look as follows:

```
../../mp.exe naip_1-1_2n_s_co011_2004_1.met
```

CURRENT ENHANCEMENTS BEING IMPLEMENTED

Sanborn experienced a few minor issues with the metadata creation which are being addressed prior to the 2005 NAIP season. One improvement needed is in the handling of exceptions. Exceptions include things such as needed files not present, handle wrong, UTM zones, etc. At present most of these types of errors are not trapped for and cause an exception in the VB software. A friendly method of detecting problem, reporting these error(s) and exiting from the application will be needed for the 2005 season.

Another item that Sanborn would like to support is the changing over and testing of ArcGIS 9.1 for the 2005 season. This transition is expected to go smoothly, and hopefully will fix a couple of minor issues.

FUTURE WORK

Additional improvements are required to fully automate the metadata creation process. Other items that must be more fully designed and considered for metadata creation are listed below:

- Integrate directly with LiftTrack.
- Look at switching to .NET for better GUIs and cleaner code base
- Look at taking the “User” out of the loop completely, except for QC
- Interface directly with the status tool at Sanborn (SanTrack)
- Performance testing
- Others

CONCLUSION

Sanborn has successfully implemented a process that used ArcGIS ArcOjects, Perl, and USGS Metadata Parser to created compliant NAIP metadata. While not without small issues, the process was an overall success and allowed the metadata to be created in a timely fashion, but more importantly was the accuracy of the metadata that was created. This task was completed with minimal difficulty due to the openness of ArcGIS and the completeness for ArcObject.

ACKNOWLEDGEMENTS

Thank you to Sanborn for giving me time and materials to work on the translator and on the support of this paper.

REFERENCES

ESRI, Exploring ArcObjects, 2001.

ESRI, System Design Strategies. March 2004.

ESRI, Shapefile Technical Description, July 1998

ESRI, ArcGIS Developer Help (both local and on-line)

USGS on-line website for metadata parser (MP)

Steven D. Straka
Senior Programmer Analyst
The Sanborn Map Company, Inc.
sstraka@sanborn.com