Integrating Digital Data Sources with DataFax

Matthew X. Economou, Network Engineer

NIH/NIAID – Office of Cyber Infrastructure and Computational Biology
Why Digital Data Capture?

- Catch errors at point of data entry
- Restrict data entry to qualified personnel
- Improve data processing efficiency
- Encourage timeliness of reporting
- Make source documents readily available to authorized personnel
Assumptions/Constraints

- Use DataFax for study management
- Function like the traditional paper model
- Validate as much as possible in the field
- Host source data locally (nothing in cloud, separate from the sponsor)
- Keep no data in the field (DLP)
- Maintain study participants’ confidentiality
- Encrypt the data transport layer
Additional Constraints

- Regulatory requirements
- Challenging operating environments
- Poor network connectivity
- Extended supply lines
- No on-site vendor support/training
General Approach
Clinical Data Capture

- No existing EHR/HIMS; CRF=source
- Android-based tablets with 3G internet
- Centralized management using Android Device Manager
- Web application (HTML5); no client needed
- Soft validation built into templates (JavaScript)
- Only certain people can input data based on their role within the study
Mobile Device Management

Activate device administrator?

Google Apps Device Policy

Click Activate to set this application as the device policy manager or click Cancel to go back.

Activating administrator will allow Device Policy to perform the following operations:

- **Erase all data**
  Erase the tablet’s data without warning by performing a factory data reset

- **Change the screen-unlock password**
  Change the screen-unlock password

- **Set password rules**
  Control the length and the characters allowed in screen-unlock passwords

- **Monitor screen-unlock attempts**
  Monitor the number of incorrect passwords entered when unlocking the screen and lock the tablet or erase all the tablet’s data if too many incorrect passwords are entered

- **Lock the screen**
  Control how and when the screen locks

- **Set password expiry**
  Control how long before lock-screen password needs to be changed

- **Set storage encryption**
  Require that stored application data be encrypted

[Activate] [Cancel]
Mi-Forms User Experience

### Mobile Forms

<table>
<thead>
<tr>
<th>Form Name</th>
<th>Description</th>
<th>Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening_12_I_073_Page1_Demographics</td>
<td>Screening, p1, Rev 21</td>
<td></td>
</tr>
<tr>
<td>Screening_12_I_073_Page2_Medical</td>
<td>Screening, p2, Rev 19</td>
<td></td>
</tr>
<tr>
<td>Screening_12_I_073_Page3_Vital</td>
<td>Screening, p3, Rev 27</td>
<td></td>
</tr>
<tr>
<td>Screening_12_I_073_Page7_Progress</td>
<td>Screening, p7, Rev 21</td>
<td></td>
</tr>
<tr>
<td>Screening_12_I_073_Page8_GPS</td>
<td>GPS, Rev 44</td>
<td></td>
</tr>
<tr>
<td>Study_12_I_073_Page7_Progress_Note</td>
<td>Study, p. 7, Rev 26</td>
<td></td>
</tr>
<tr>
<td>Study_12_I_073_Page8_AEs</td>
<td>Study, p. 8, Rev 33</td>
<td></td>
</tr>
<tr>
<td>Study_12_I_073_Page9_ConMeds</td>
<td>Study, p. 9, Rev 41</td>
<td></td>
</tr>
</tbody>
</table>
Mi-Forms User Experience

![Screening_12_I_073_Page2_Medical](image)

### Subject/Visit

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject ID Prefix</td>
<td>12073</td>
</tr>
<tr>
<td>Subject ID #</td>
<td></td>
</tr>
<tr>
<td>Site #</td>
<td>4020</td>
</tr>
<tr>
<td>Expected range</td>
<td>01/09/2012 - today (dd/mm/yyyy)</td>
</tr>
</tbody>
</table>

### Medical History

- If Abnormal, comment: **No value selected**
- 1. Systemic Symptoms, e.g. fever, diabetes:
  - Yes
  - No
### Soft Validation

#### Subject/Visit

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject ID Prefix</td>
<td>12073</td>
</tr>
<tr>
<td>Subject ID #</td>
<td>522</td>
</tr>
<tr>
<td>Site #</td>
<td>4020</td>
</tr>
<tr>
<td>Expected range</td>
<td>01/09/2012 - today (dd/mm/yyyy)</td>
</tr>
<tr>
<td>Visit Date</td>
<td>15-08-2014</td>
</tr>
</tbody>
</table>

#### Medical History

1. Systemic Symptoms, e.g. fever, diabetes

- If Abnormal, comment
  - No value selected
  - Yes
  - No
Data Capture Toolchain

- Mi-Forms web application
- Microsoft SQL Server back-end database
- Mi-Forms graphical template design tools
- Page-level role-based access controls (within Mi-Forms)
- Built-in audit logging
- Custom data adapter (VB.NET):
  - Exports data via Mi-Forms.NET API
  - Transforms into DFimport text format
  - Also outputs PDF
Template Design
Role-Based Access Controls
# Role-Based Access Controls

## Edit Group

### Status
- Active

### Group Name
- Clinician

### Privileges
- Admin
- Form Filler
- Publisher
- Template Filler
- User

### Form Templates
- Allowed
  - Screening_12_1_073_Page1_Demographics
  - Screening_12_1_073_Page2_Medical
  - Screening_12_1_073_Page3_Vital
  - Screening_12_1_073_Page4_TB
  - Screening_12_1_073_Page7_Progress
  - Study_12_1_073_Page10_EOS
  - Study_12_1_073_Page2_Stool_Day0
  - Study_12_1_073_Page4_Stool_Day0
  - Study_12_1_073_Page7_Progress_Note
  - Study_12_1_073_Page8_AEs
  - Study_12_1_073_Page9_CompMeds

- Not Allowed
  - Screening_12_1_073_Page5_Blood
  - Screening_12_1_073_Page6_Hematology
  - Screening_12_1_073_Page8_GPS
  - Study_12_1_073_Page1_Eligibility_Day0
  - Study_12_1_073_Page3_CD8_Day0
  - Study_12_1_073_Page5_6_Month6

### Users
- In Group
  - knewell
  - kwatkins
  - mholdsworth
  - pbalakrishnan
  - smuthusamy
  - smuthusamy2

- Not In Group
  - administrator
  - jganesh
  - jganesh2
  - jgeorge
  - jgeorge2
  - pub
  - sbabu
  - sbabu2
Audit Logging

<table>
<thead>
<tr>
<th>Subject/Visit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject ID #</td>
<td>1234</td>
</tr>
<tr>
<td>Site #</td>
<td>4020</td>
</tr>
<tr>
<td>Visit Date</td>
<td>09-07-2012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Age (years)</td>
<td>32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>09-06-2012</td>
<td>20:22:45</td>
<td>administrator</td>
</tr>
<tr>
<td>09-06-2012</td>
<td>12:14:48</td>
<td>testuser</td>
</tr>
<tr>
<td>09-06-2012</td>
<td>11:27:12</td>
<td>testuser</td>
</tr>
</tbody>
</table>
Data Export/Transformation

' Two usage scenarios: page on the form, and hard-coded page number (passed as prefix)
Public Function LinePrefixWithPage(ByVal PlateID As String, ByVal PagePrefix As String) As String
    ' Note: the SubjectID is taken from the first page. DataFax expects that it's uniform throughout the form.
    LinePrefixWithPage = "1\[0000/00000000-" & StudyID & "|" & PlateID & "|" & PagePrefix & ".PageNum.PadLeft(3 - PagePrefix.Length, "0")
End Function

'DataFax line suffix
'RecStatus CreationDate LastModDate
Private Function LineSuffix() As String
    LineSuffix = "1|" & DirectCast(_fbo._CreationDate, Hidden).Value & "|" & Now.ToString("yy") & "|" & Now.ToString("MM") & "|" & Now.ToString("dd")
End Function

'Terminates each field with a | for better readability.
'When passing strings, make sure they're neither started nor trailed by |
Public Sub Append(ByVal a As Object())
    Dim i As Integer
    Dim o As Object
    For i = LBound(a) To UBound(a)
        o = a(i)
        Dim v As String = ""
        If TypeOf o Is String Then
            v = DirectCast(o, String)
        ElseIf TypeOf o Is FormControl Then
            v = (DirectCast(o, FormControl)).Value.Trim() 'Tested on TTextField, CheckBoxGroup, FreeForm, InkX (assuming text entry)
            'Extra treatment for numeric controls with fractions
            If IsFractional(o) Then v = v.TrimEnd(Chr(13)).Replace(Chr(10), "")
        End If
        If v.IndexOf(Chr(13)) >= 0 Then v = "" & v.Replace(Chr(13), "")
        If v.IndexOf(Chr(9)) >= 0 Then v = "" & v.Replace(Chr(9), "")
        If v.IndexOf(Chr(8)) >= 0 Then v = "" & v.Replace(Chr(8), "")
        s = s & v & "|
    Next
End Sub
1|1|0000/0000000|220|002|1|120730522|4
020|15/10/2013|2||2||2||2||2||2||MS|15/10
/2013|1|13/10/21 10:57:33|13/10/21
10:57:33|
Uploading Data into Studies

- Data adapter runs daily
- rsync over SSH file transfer
- dos2unix for EOL conversion
- DFimport.rpc loads data into study
- No image associated (like all EDC studies)
- Normal DataFax workflow from there
Data Transfer Process

```bash
move the files into a temporary directory ahead of the file
# transfer
ssh s-msrc2dfax@source.trecchennai.in 'cd /cygdrive/e/DataTransfer/Upload; mkdir ${MODE}${TIMESTAMP} ; /usr/bin/find Ready/${MODE} -maxdepth 1 -mindepth 1 -exec mv '{}' ./${MODE}${TIMESTAMP}/ ;'

# download the files from the temp directory
rsync -rtv --chmod=ugo+rX,Fu+x,g+o+r s-msrc2dfax@source.trecchennai.in:/cygdrive/e/DataTransfer/Upload $MODE$TIMESTAMP ~

# move the files to the "Done" directory after the transfer completes
# (create the folder if missing)
ssh s-msrc2dfax@source.trecchennai.in 'cd /cygdrive/e/DataTransfer/Upload; /usr/bin/find $MODE$TIMESTAMP -maxdepth 1 -mindepth 1 -exec mv '{}' /Done/$MODE/ ; ; rmdir $MODE$TIMESTAMP'

for i in `find $MODE$TIMESTAMP -name '*.txt' -type f -print | sort`; do
  # convert to Unix line endings
dos2unix $i

  # strip any non-ASCII characters; set LANG to work around error "invalid collation character"
  env LANG=C sed -i"" -e 's/\x80-\xFF\//g' $i

  # import the records into Datafax
  DFimport.rpcs -v -R -m $[STUDY] $i 2>&1 | tee results.out

  # check for any errors or warnings; email them to the data manager(s)
  if [ -z "egrep '0 failed' results.out" -o -z "egrep '0 warning' results.out" ]; then
    mail -s "$[MODE] DFimport errors/warnings on file `basename $i'" -b datafax@niaid.nih.gov < results.out
  fi

  # archive the file
  mv $i ~/archive
done

# archive any remaining files (e.g., PDFs)
find ~/$MODE$TIMESTAMP -maxdepth 1 -mindepth 1 -exec mv '{}' ~/archive \;
rmdir ~/$MODE$TIMESTAMP
```

---

m2df.sh   Bot (21.0)   (Shell-script[sh] Filladapt) 4:58PM 1.68
Data Transfer Logging

+ MODE=Prod
+ STUDY=220
+ PATH=/opt/datafax/bin:/usr/kerberos/bin:/usr/local/bin:/bin:/usr/bin
+ export PATH
++ date -u +%Y%m%d%H%M
+ TIMESTAMP=201409151730
+ '['=''-d /home/msrc2dfax/archive ']''
+ ssh s-msrc2dfax@msource.trcchennai.in 'cd /cygdrive/e/DataTransfer/Upload; mkdir Prod201409151730; /usr/bin/f
+ rsync -rtv --chmod=Dgo+rX,Fu-x,g0+r s-msrc2dfax@msource.trcchennai.in:/cygdrive/e/DataTransfer/Upload/Prod201
receiving incremental file list
Prod201409151730/
Prod201409151730/DataImportFile_120730569_20140915_121332.txt
Prod201409151730/DataImportFile_120730917_20140915_105610.txt
Prod201409151730/DataImportFile_120730965_20140915_105539.txt
Prod201409151730/DataImportFile_120731014_20140915_105629.txt
Prod201409151730/DataImportFile_120731074_20140915_105706.txt
Prod201409151730/DataImportFile_120731099_20140915_121422.txt
Prod201409151730/DataImportFile_120731105_20140915_121451.txt
Prod201409151730/DataImportFile_120731129_20140915_105725.txt

sent 167 bytes  received 1971 bytes  328.92 bytes/sec
total size is 1285  speedup is 0.60
+ ssh s-msrc2dfax@msource.trcchennai.in 'cd /cygdrive/e/DataTransfer/Upload; /usr/bin/find Prod201409151730 -ma
++ find Prod201409151730 -name '/*.txt' -type f -print
++ sort
+ for i in `find $(MODE)$({TIMESTAMP}) -name '/*.txt'` -type f -print | sort`
+ dos2unix Prod201409151730/DataImportFile_120730569_20140915_121332.txt
dos2unix: converting file Prod201409151730/DataImportFile_120730569_20140915_121332.txt to UNIX format ...
+ env LANG=C sed "-i -e 's/\[\x80\-\xff]/\g' Prod201409151730/DataImportFile_120730569_20140915_121332.txt
tee results.out
+ DFimport.rpc -v -R -m 220 Prod201409151730/DataImportFile_120730569_20140915_121332.txt

1 records input: 1 imported (0 warnings), 0 failed with errors
+ egrep '0 failed' results.out
++ egrep '0 warning' results.out
+ '[' -z '1 records input: 1 imported (0 warnings), 0 failed with errors' -o -z '1 records input: 1 imported (0
+ mv Prod201409151730/DataImportFile_120730569_20140915_121332.txt /home/msrc2dfax/archive
### Data sans Fax

<table>
<thead>
<tr>
<th>Subject ID #</th>
<th>120730522</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit Date</td>
<td>15/10/2013</td>
</tr>
</tbody>
</table>

#### SCREENING

**MEDICAL HISTORY**

<table>
<thead>
<tr>
<th>Body System</th>
<th>Finding</th>
<th>If Yes, Comment:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Systemic Symptoms (e.g. fever, diabetes)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2. Pulmonary (e.g. cough, hemoptysis)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>3. Cardiovascular (e.g. chest pain, hypertension)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>4. Lymphatic (e.g. lymph node swellings, hydrocele, lymphedema)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>5. Any other major illness</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

1| 1| 0000/0000000| 220| 002| 1| 120730522| 4020| 15/10/2013| 1| 13/10/21 10:57:33| 13/10/21 10:57:33|
Risks and Mitigations

- Confidentiality/integrity/availability
  - Mobile device management (MDM)
  - Transport-layer encryption (HTTPS, SSH)
  - Data at rest in the investigator’s data center (redundant AC/power/Internet, backups/DRP/COOP, etc.)
- Only authorized personnel
  - Access controlled by Mi-Forms server
- Protocol changes
  - CRF updates automatically pushed OTA
Risks and Mitigations

- Log all data access (for audit later)
  - Activity logs in Mi-Forms
  - File transfer logs
- CRF development overhead
  - Trained staff on both Mi-Forms and DataFax form development
- Poor network connectivity
  - Primary network connection via 3G
  - Backup network connection via WiFi mobile routers
Risks and Mitigations

- No image files accompanying data
  - Correction history still accessible via the audit log
- What if something goes wrong?
  - Revert to paper if necessary
  - All exported data automatically archived
  - Automatically outputs to PDF archive
  - Can also output to CDISC ODM if needed
Key Features and Benefits

- Immediate review of data inputs (soft validation)
- Centralized control over form contents (OTA updates)
- Page access restricted to qualified personnel (RBAC)
- Automatic Mi-Forms-to-DataFax ETL
- Improved timing/throughput (e.g., for real-time safety monitoring)
ACKNOWLEDGEMENTS

- NIAID/Office of Cyber Infrastructure and Computational Biology
  - Michael Tartakovsky
  - Alex Rosenthal
  - Ruma Das
  - Chris Whalen
  - Kevin Newell
  - Michael Holdsworth
  - Michael Duvenhage
  - Jiwen Sun
  - Cynthia Lassnoff
  - Brian Moyer
  - Arul Mozhi
  - Vsevolod Alekseyev

- RCHSPB
  - Susan Vogel
  - Lisa Hoopengardner
  - Katie Watkins

- NIAID Intramural
  - Subbaraman Babu
  - Thomas Nutman

- NIRT
  - Soumya Swaminathan
  - Prabbu Balakrishnan

- DF/Net Research

- Mi-Corporation
Thank You!