

---

## DataFax Archiver

– a Web-based Tool for DataFax Study Close-out and Archiving

Hanming Tu  
Premier Research Group plc  
Centre Square West, 1500 Market Street, Suite 3500  
Philadelphia, PA 19102  
Email: [Hanming.Tu@premier-research.com](mailto:Hanming.Tu@premier-research.com)  
Phone: 215-282-5526



## Content

---

- ❖ Data Transferring Project
- ❖ The Tools
- ❖ The Issues
- ❖ YAML
- ❖ DFarchiver As a DFOSP



## Data Transferring Project

---

- ❖ Alza project: transfer 6 studies conducted during 1997 and 2000
- ❖ Tasks:
  - Restore the studies (all older than 3 years)
  - Generate PDF files for all CRFs
  - Package each study individually
  - QC the final files
  - Burn onto CDs



## Archived Study Files

---

- ❖ DataFax Data
  - All DataFax files
  - Tar creation log
  - Tar extraction log
- ❖ Integrity Check Reports
  - DF\_ICimages: Verify image references against data records.
  - DF\_ICkeys: Check key fields (id,visit,plate) and patient initials with primary records only (-R primary).
  - DF\_ICqcs: Check QC notes for consistency with data records.
  - DF\_ICrecords: Check data records for structural inconsistencies.
  - DF\_ICschema: Check database schema for common errors and inconsistencies.
  - DF\_ICvisitdates: List QC export errors from the last execution of DF\_XXkeys.
  - DF\_ICvisitmap: Report inconsistencies and errors in the study visit map.



## Archived Study Files

---

### ❖ Patient PDF Files

- One PDF per patient
- One Data file per patient

### ❖ Program Files:

tar.exe, gzip.exe, and WIN32GNU.DLL

### ❖ Example

Study 071



## Example Screens

---

### Study 071 - Study Title Concealed

Here is a list of all the index files for this study:

- [/DF/S071c.txt](#): a list of all DataFax files (tar creation log).
- [/DF/S071t.txt](#): a list of all DataFax files (tar listing log).
- [/DF/S071.tar.gz](#): the archived and compressed DataFax files (size 18,172 kb). You need to use `gzip` and `tar` or `WinZip` to expand the file. I have included the DOS version of the `gzip.exe` and `tar.exe` with this delivery. You can also use UNIX version of `gzip` or `tar` commands.
- [/ICs/index.html](#): Reports for DataFax integrity checks.
- [/PDFs/index.html](#): DataFax CRFs in PDF format.

---

© Copyright Premier Research, a Division of CRC Development, Ltd.  
30 South 19th Street, Philadelphia, PA 19103



# Example Screens

## Study 071 - Title Concealed

Here is a list of integrity checks being done for this study:

- DF\_ICimages: Verify image references against data records.
- DF\_ICkeys: Check key fields (id,visit,plate) and patient initials with primary records only (-R primary).
- DF\_ICqcs: Check QC notes for consistency with data records.
- DF\_ICrecords: Check data records for structural inconsistencies.
- DF\_ICschema: Check database schema for common errors and inconsistencies.
- DF\_ICvisitdates: List QC export errors from the last execution of DF\_XXkeys.
- DF\_ICvisitmap: Report inconsistencies and errors in the study visit map.

**Note:** All the reports were run with defaults except DF\_ICkeys. DF\_ICkeys was run with "-R.primary".

© Copyright Premier Research, a Division of CRC Development, Ltd.  
30 South 17th Street, Philadelphia, PA 19103



# Example Screens

## CDI contains:

```

Started at Thu Sep  9 16:59:44 EDT 2004
S071 0108001.txt, S071 0108001.pdf (Bytes: 376284; Created at: Sep  9 16:59; Records: 26)
S071 0108002.txt, S071 0108002.pdf (Bytes: 326786; Created at: Sep  9 16:59; Records: 24)
S071 0108003.txt, S071 0108003.pdf (Bytes: 410620; Created at: Sep  9 16:59; Records: 28)
S071 0108004.txt, S071 0108004.pdf (Bytes: 371772; Created at: Sep  9 16:59; Records: 26)
S071 0108005.txt, S071 0108005.pdf (Bytes: 194056; Created at: Sep  9 16:59; Records: 15)
S071 0108006.txt, S071 0108006.pdf (Bytes: 342016; Created at: Sep  9 16:59; Records: 24)
S071 0108007.txt, S071 0108007.pdf (Bytes: 385262; Created at: Sep  9 16:59; Records: 27)
S071 0108008.txt, S071 0108008.pdf (Bytes: 272825; Created at: Sep  9 16:59; Records: 19)
S071 0108009.txt, S071 0108009.pdf (Bytes: 374661; Created at: Sep  9 16:59; Records: 26)
S071 0108010.txt, S071 0108010.pdf (Bytes: 342478; Created at: Sep  9 16:59; Records: 25)
S071 0208001.txt, S071 0208001.pdf (Bytes: 593192; Created at: Sep  9 16:59; Records: 35)
S071 0208002.txt, S071 0208002.pdf (Bytes: 409048; Created at: Sep  9 16:59; Records: 28)
S071 0408001.txt, S071 0408001.pdf (Bytes: 373315; Created at: Sep  9 16:59; Records: 26)
S071 0408002.txt, S071 0408002.pdf (Bytes: 448672; Created at: Sep  9 16:59; Records: 30)
S071 0408003.txt, S071 0408003.pdf (Bytes: 412162; Created at: Sep  9 16:59; Records: 28)
S071 0408004.txt, S071 0408004.pdf (Bytes: 325265; Created at: Sep  9 16:59; Records: 22)
S071 0408005.txt, S071 0408005.pdf (Bytes: 329374; Created at: Sep  9 16:59; Records: 23)
S071 0408006.txt, S071 0408006.pdf (Bytes: 307664; Created at: Sep  9 16:59; Records: 23)
S071 0408007.txt, S071 0408007.pdf (Bytes: 395221; Created at: Sep  9 17:00; Records: 27)
S071 0408008.txt, S071 0408008.pdf (Bytes: 364865; Created at: Sep  9 17:00; Records: 26)
S071 0408009.txt, S071 0408009.pdf (Bytes: 381804; Created at: Sep  9 17:00; Records: 27)
S071 0408010.txt, S071 0408010.pdf (Bytes: 207382; Created at: Sep  9 17:00; Records: 15)
Total records - 550 and bytes - 7944724
Ended at Thu Sep  9 16:59:59 EDT 2004

```



## PDF Tool - DF\_ATpdf

### ❖ DF\_ATpdf has the following features:

- Integrated into DataFax report: run within DataFax, in command line, and in web reporter.
- Inherent all fine selection from DFexport.rpc
- Use DFpdf as the PDF engine
- Divide PDF files into a CD or DVD full and linked all the files with a HTML index file
- Transfer PDF files to a burn station
- Advanced options for study federation



## DF\_ATpdf Web Interface

**DataFax Web Reporter**

Report_source:	<input checked="" type="radio"/> DataFax Generic <input type="radio"/> Study Specific	<input checked="" type="radio"/> Local <input type="radio"/> Delayed <input type="radio"/> Real time <input type="checkbox"/> Verbose <input checked="" type="checkbox"/> Format output
Select studies:	001 - Mellin 1092-96-01	<input type="checkbox"/> View Codes <input type="checkbox"/> Save
Select a report:	DF_ATpdf - Create PDF with bookmarks (formerly DF_runDFpdf)	
Set options:	-a <archive_directory d default> ..PDF directory -S primary -O P -a d -g	
Actions:	<input type="button" value="Run Report"/>	<input type="button" value="Explain"/> <input type="button" value="Refresh"/> <input type="button" value="Reset"/>

DF\_ATpdf - Create PDF with bookmarks (formerly DF\_runDFpdf)

```
DF_ATpdf [-P #-#] [-v #-#] [-I #-#] [-V #-#] [-C #] [-M #] [-S #]
          [-o file] [-p pagedir] [-f filemap] [-c title] [-s #]
create PDF file using output from DFexport.rpc as inpput.
OPTIONS:
-h          display this help information.
.
```



## PDF Tool Parameters

### ❖ Record Selection options (DFexport.rpc)

- P <plate\_list> a list of plates in the free range format (FRF). The default is to include all the plates.
- v <valid\_list> a list of validation levels in the FRF. By default, records at all validation levels are exported.
- I <id\_list> patient identifier. If this argument is specified, only those records which have a patient identifier matching the selection criteria are output.
- V <visit\_list> visit/sequence number.
- C <create\_list> creation date. Only to the creation date of data records and not QC notes.
- M <modify\_list> modification date. Only to the modification date of data records and not QC notes.
- S <status> record status. The recognized status keywords are: "new clean dirty error CLEAN DIRTY ERROR lost all delete primary secondary" and any combination can be given as a quoted string. The default is "all", which exports all primary, secondary and lost (but not pending delete) data records, or all QC notes if the export plate is 511.



## PDF Tool Parameters

### ❖ Customized Options:

- h display this help information.
- d <0 510 511> additional plate number separated by space: 0 - DFin.dat, 510 - for reason for data change records, 511 for quality control notes.
- g generate index file for each CD folder
- x exclude text files from index page and file transfer
- s study numbers, required.

For study on a remote server, you can specify as the following:  
SN:SVR:DataFax\_DIR.

To add a secondary study, use this input format:

SN:p\_SVR:p\_DFDIR,s\_SN:s\_SVR:s\_DFDIR

For example, to combine study 82 and a remote study 83 into PDF creation using '-s 82,83:df-svr2:/opt/datafax'



## PDF Tool Parameters

---

### ❖ Output Options (DFpdf)

- o outfile: the output file that will contain the resulting PDF document.
- p pagedir: an alternative root directory at which CRF images can be located.
- f filemap: an alternative file location for the study filemap.
- m visitmap: an alternative file location for the study visitmap.
- t title: the title to appear at the top of the bookmark list.



## The Issues

---

- ❖ Time consuming: although the PDF tool helped in generating PDF files but all the other files were created manually
- ❖ Standard procedure (SOP): we need to perform the similar procedure many times in the lifespan of a study or even in the post-study: pre-audit, post-audit, pre-closeout, post-closeout, and archiving
- ❖ Manageability: How to manage the scattered files? Can we have a study portal?
- ❖ Interoperability: what if companies do not have DataFax and even SAS? How can other system to use the data?



## YAML – Is It An Answer?

---

### ❖ What Is YAML?

- New document format used mainly for configuration and data output for further processing.
- A human-friendly, cross language, Unicode based data serialization language
- Designed around the common native data structures of agile programming languages.
- It is broadly useful for programming needs ranging from configuration files to Internet messaging to object persistence to data auditing



## Why YAML?

---

- ❖ YAML is readable for people
- ❖ YAML is editable
- ❖ YAML is multilingual
- ❖ YAML is taint safe
- ❖ YAML is full featured
- ❖ YAML is extensible



## Text and XML Files

---

### ❖ Text File

```

FirstName|LastName|DOB|Sex|BP|Weight
John|Smith|23Dec1962|M|120/70|180 lbs
Tina|McFong|15Apr1982|M|114/68|125 lbs

```

### ❖ XML File

```

<demo><FirstName>John</FirstName><LastName>Smith</L
astName><DOB>23Dec1962</DOB><Sex>M</Sex><BP>120/7
0</BP><Weight>180 lbs</Weight></demo>
<demo><FirstName>Tina</FirstName><LastName>McFong</
LastName><DOB>15Apr1982</DOB><Sex>M</Sex><BP>114/
68</BP><Weight>125 lbs</Weight>
</demo>

```



## YML File Example

---

### ❖ YML File:

```

---
demo:
  - '': ~
    FirstName: John
    LastName: Smith
    DOB: 23Dec1962
    Sex: M
    BP: 120/70
    Weight: 180 lbs
  - '': ~
    FirstName: Tina
    LastName: McFong
    DOB: 15Apr1982
    Sex: M
    BP: 114/68
    Weight: 125 lbs
...

```



## DFarchiver – A DataFax Open Source Project

---

- ❖ Purpose: develop a web-based interface to help in DataFax closeout and archiving.
- ❖ Requirements:
  - Preserves DataFax native files
  - Use XML and YML as main viewable formats
  - Use CDISC standards whenever it is possible
  - Generates other files such as HTML, PDF, CSV when necessary



## Archiving Approach

---

- ❖ Proposed Approach:
  - Closing a study
  - Documenting files
  - Indexing files



## Closing a Study

---

### ❖ Pre-database lock

- Define a database lock date
- Agree on policy for handling new faxes
  - Print and remove new faxes from the new queue or
  - Continue validation new faxes by one user
- Perform pre-database lock checks
  1. All records have been validated: `DFstatus -c -s 253`
  2. All CRFs have been received (no missing pages or overdue visits): `DF_PTmissing 253`
  3. All QC notes have been resolved: `DF_ICqcs 253` (run `DF_QCupdate`, `DF_QCreport`, `DF_QCprint` if there is any issue)
  4. All required data is clean, with no error status and reviewed according to study SOPs: `DF_ICrecords 253 -d`
  5. No errors in all other consistency checks: `DF_ICvisitdates`, `DF_ICkeys`, `DF_ICimages`, `DF_ICvisitmap`, `DF_ICschema` (the last two just for documentation purpose)



## Closing a Study

---

### ❖ Lock the database

- Compact plates: `DFsortplate.rpc -f -s 253`
- Print current study permissions: `DFuserPerms -l :253`
- Remove user permissions:
 

```
foreach u `DFuserPerm -l :253 | cut -d ` ` -f1`
  DFuserPerms -m ${u}:253:-s0
end
```
- Bring all primary records to validation level 5 or whatever level that your company defines
- Remove write permission to lib and its contents
- Monitor new record queue and audit the DB



## Closing a study

---

### ❖ Post-database lock

#### ➤ Re-check DB










1. All CRFs have been received (no missing pages or overdue visits): `DF_PTmissing 253`
2. All QC notes have been resolved: `DF_ICqcs 253` (run `DF_QCupdate`, `DF_QCreport`, `DF_QCprint` if there is any issue)
3. All required data is clean, with no error status and reviewed according to study SOPs: `DF_ICrecords 253 -d`
4. No errors in all other consistency checks: `DF_ICvisitdates`, `DF_ICkeys`, `DF_ICimages`, `DF_ICvisitmap`, `DF_ICschema` (the last two just for documentation purpose)
5. Bring all primary records to validation level 7



## Documenting Files

---

### ❖ Files to be documented:

-  DataFax users
-  DataFax study configurations
-  DataFax data files
-  DataFax image files
-  DataFax audit files
-  SAS programs
-  Final client data sets
-  Project management documents
-  Index files for various components



## Indexing Files

---

- ❖ Index files will be created for each type of documents
- ❖ Viewable documents will be linked
- ❖ Navigation links will be added for each index pages and for some viewable documents



## The Web Interface

---

**PREMIER RESEARCH DataFax Archiver**

Select Study  ==>

Options  DataFax Generic  Study Specific

Action

- Select
- Check configurations
- Pre-database lock
- Lock the database
- Post-database lock
- Convert data files
- Tar DataFax files
- Generate index files



## Perl Modules

---

### ❖ The modules that I have uploaded to CPAN

- DataFax
- DataFax::StudyDB
- DataFax::StudySubs
- CGI::AppBuilder and its six sub modules
  - CGI::AppBuilder::Config
  - CGI::AppBuilder::Message
  - CGI::AppBuilder::Form
  - CGI::AppBuilder::Table
  - CGI::AppBuilder::Header
  - CGI::AppBuilder::Frame
- Oracle::Loader, Oracle::Trigger, Oracle::DML, etc.

### ❖ All of them are object-oriented Perl (OOP)



## Perl Modules

---

### ❖ The planned modules:

- DataFax::Lib
- DataFax::Reports
- DataFax::Data
- DataFax::CDISC
- DataFax::Validation
- Etc.
- ➔DataFax::Bundle

### ❖ How to install and import modules

- Perl -MCPAN -e "install DataFax"
- Perl -MCPAN -e "install DataFax::Bundle"
- Use DataFax ':all'



## Join DFOST

---

- ❖ It is a big task and hope many would join the DataFax Open Source Team
- ❖ I have created a project at SourceForge:  
<http://sourceforge.net/projects/dfosp>
- ❖ Some Perl modules are published in CPAN.



## Q & A

---

