

---

# Enable CDISC With DataFax

Hanming Tu  
SCP Communications, Inc.

## Content

---

- **The problems:**
  - Common Issues
  - Common Requirements
- **General Solution:**
  - Common Standards
  - CDISC and ODM
- **Our Solution:**
  - System: DataFax with ICTMS
  - Technology: XML with CDISC
  - Implementation: Web with ETL Engine
- **Conclusion**

## Common Issue – Data

---

- Clinical data comes from diverse sources:
  - CRFs on paper or in file
  - Safety Laboratories
  - ECG machines
  - Electronic data capture and patient diary systems
- Clinical data comes in various data formats:
  - DataFax, SAS, Word, Excel, etc;
  - ASCII vs. Binary

## Common Issue - System

---

- Incompatible systems
  - Data are typically passed among multiple computer systems during collecting, cleansing, analyzing, and submitting.
- Different means for data access:
  - DataFax Reports, Oracle Discovery, Brio Query, Crystal Report, etc.
  - Paper vs. electronic

## Common Issue - Organization

---

- Multiple organizations
  - Contract Research Organization
  - Central labs and ECG core labs
  - EDC vendors
  - Merged companies and/or affiliates
  - Pharmaceutical Sponsors

## Common Requirement - Data

---

- Complex requirements for data integrity.
  - Protect data and its linkage information against deliberate falsification as well as accidental loss.
  - Link electronic clinical records to electronic signatures.
  - Maintain a secure, time-stamped audit trail.
  - Access to metadata and presentation information.

## Common Requirement – Audit

---

- Rigorous audit and review requirements.
  - Must generate accurate and complete records in electronic form suitable for agency inspection, review, and copying.
  - Must support drill down to verify data integrity.

## Common Requirement - Retention

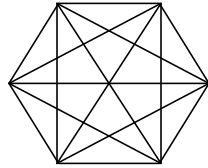
---

- Extended data retention requirement.
  - Clinical data and audit trail must be retained and protected to enable accurate and ready retrieval throughout the retention period.
  - The longer the retention, the more costly it is.

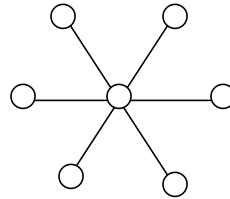
## Core of the Issues

---

- Lack of standards
- Too many standards: it is the same as lack of them



No standard



With standard

Systems	3	4	5	6	7	8	9	10
No Std	3	6	10	15	21	28	37	46
With Std	3	4	5	6	7	8	9	10

## General Solution – SDO and FDA

---

- **SDO and Harmonization:**
  - HL7: Health Level 7
  - ICH: The International Conference on Harmonization
  - ISO: International Standard Organization
  - W3C: World Wide Web Consortium
  - CDISC: Clinical Data Interchange Standards Consortium
- **FDA and Legalization:**
  - HIPAA: The Health Insurance Portability and Accountability Act
  - FDA: Which standard FDA will use?

SDO = Standards Developing Organizations

## General Solution – Adopting a Standard

---

- **Our searching and struggling:**
  - Resource Planning and Management for Clinical Data Management using DataFax, Oracle and Web server (DFUG 1999, Lake Louise)
  - Bridging DataFax with Oracle
  - DataFax Web Reporter (DFUG 2000, Whistler, BC)
  - Web Portal: From 'D' to 'E' (DFUG 2000, Whistler, BC)
- **Better to use a common standard – CDISC**

## CDISC - Mission

---

- Clinical Data Interchange Standards Consortium is an open, multidisciplinary, non-profit organization committed to the development of worldwide industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trials data and metadata for medical and biopharmaceutical product development.
- The CDISC mission is to lead the development of global, vendor-neutral, platform-independent standards to improve data quality and accelerate product development in our industry.

## CDISC – Principles

---

- **Scientific:** to lead the development of standard data models to improve process efficiency, to support the scientific nature of clinical research, and allow flexibility in scientific content in regulatory submission.
- **Independent:** to maintain a global, multidisciplinary, cross-functional composition for CDISC and its working groups and to keep its standards and models independent from specific implementation strategy and platform.
- **Open:** to encourage maximum sharing of information and minimum duplication.
- **Educational:** to provide educational programs on CDISC standards, models, values and benefits.

## CDISC – Models and Standards

---

- **Operational data models**
  - ODM: operational data model, including study data, audit trail, and metadata
  - LAB: lab data content model
- **Submission data models**
  - SDM: submission data model
  - ADaM: analysis data model
  - SMM: submission metadata model

## CDISC – Current Status

---

- **ODM 1.1.**
  - Final version done in April 2002
  - Testing is completed
- **Lab 1.0.**
  - Comments from public review currently being addressed
  - Final version is posted on 11/6/2002
- **SDM 2.0.**
  - Version 1.0 published in October 2000, version 1.1 in June 2001
  - Current version completed in November 2001
  - Participation in FDA pilot for patient profile viewer.
- **ADaM.**
  - Guidelines and example completed in November 2001;
  - Other examples are in progress.

## ODM – Technology: XML

---

- ODM is a XML-based, CDISC-compliant, and open data model.
- XML is well suited to clinical data applications
  - Metadata support provides needed flexibility
  - ASCII based is well suited to long term use
- Increasing acceptance by FDA and ICH
  - New electronic submissions guidance
  - Common technical document

## ODM – Compliance

---

- ODM must satisfy correctness rules in CDISC standards, both syntactic and semantic.
  - Syntactic – well formed, properly referenced XML file
  - Semantic – documented in [odm1-1-0.html](#)
- ODM contains administrative data, signature and audit structures - providing compliance with 21 CFR part 11.
- A receiving system must read and interpret ODM files correctly and may selectively ignore non-required or extended information.

## ODM – Openness

---

- ODM is an open standard model and still evolving.
- Current ODM allows extensions but the extensions may be ignored by receiving systems.
- Standards vs. Creativity

## ODM - Content

---

- Captures metadata, administrative data, and reference data associated with patient clinical data.
- Events structure in metadata and clinical data reflects CRF/eCRF data.
- Administrative data, signature and audit structures provide compliance with 21 CFR part 11.
- Includes many ease of use features to support a variety of use case scenarios.

## ODM - Limitations

---

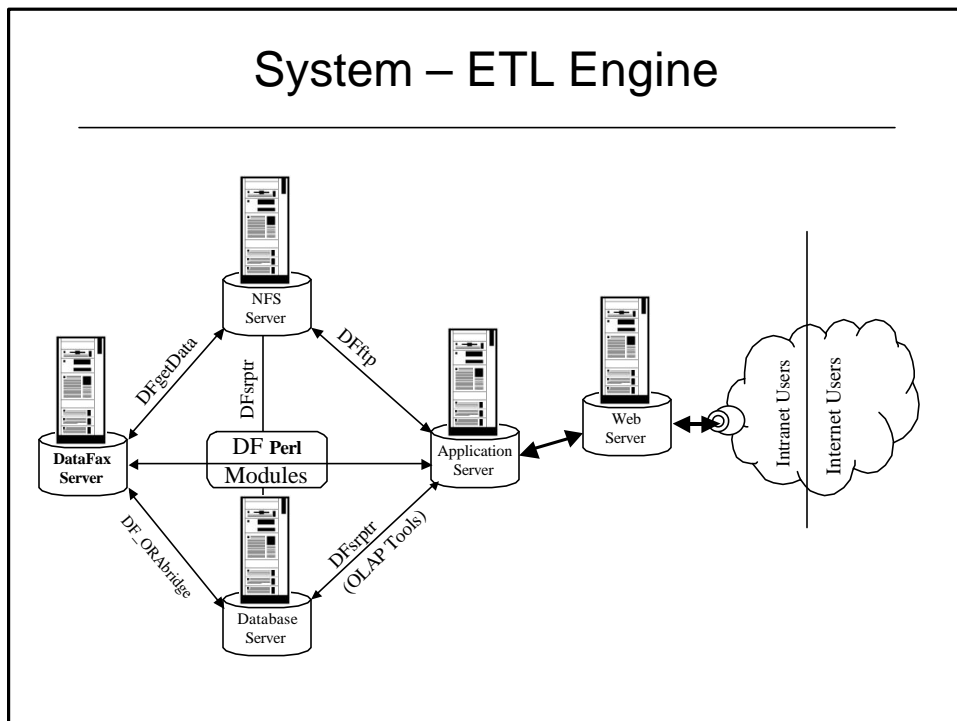
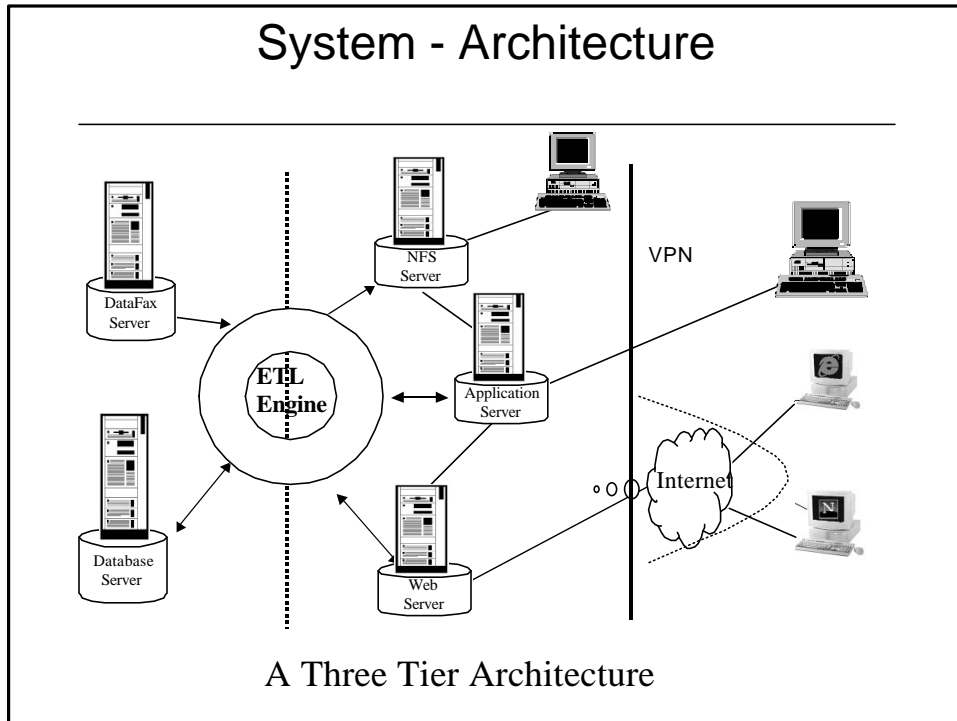
- ODM 1.1 only uses XML DTD, next version will use XML schema
- ODM does not include clinical trial project management (our PPM)
- ODM does not include user permissions or access levels for forms or variables.
- ODM does not include lab records, which are in LAB standard. We will have to join lab records in submission data

## System - DataFax

- Standards need systems to support them.
- DataFax is a proven and effective system for automatically capturing data through fax.
- It is easy to enable it to export data in CDISC format.

## System - ICTMS





## Technology – XML and Perl

---

- As we reasoned before, we have chosen XML as our technology and CDISC as our standard.
- We have chosen Perl as ETL language
  - Widely used language
  - Web conscious language
  - Easy to port to different platforms
  - Good at processing text
  - Object oriented implementation
  - Can be very fast if used in mod\_perl with Apache
  - Many shared modules
  - It is FREE!

## Technology – Perl XML Modules

---

- We use many Perl modules obtained from Comprehensive Perl Archive Network (CPAN: <http://search.cpan.org/>):
  - XML::SAX
  - XML::Parser
  - XML::Generator
  - XML::Simple
- The above modules use Expat as parsing engine:
  - Expat - a fast, non-validating, stream-oriented XML parsing library written in C.
  - Expat can be obtained from <http://www.sunfreeware.com/>.

## Implementation – Perl DataFax Modules

---

- We developed and planned to develop the following modules to form our ETL engine:
  - Fax::DataFax – a top module to enable DataFax awareness through configuration file (DataFax.ini).
  - Fax::DataFax::CDISC – enable CDISC awareness
  - Fax::DataFax::ODM – ETL class for Operational Data Model
  - FAX::DataFax::Lab – ETL class for Laboratory data model
  - Fax::DataFax::SDM – ETL class for Submission Data Model
  - Fax::DataFax::ADaM – ETL class for Analysis Dataset Model

## Implementation – Supporting Perl Modules

---

- We developed many supporting modules:
  - Data::Loader – class for loading parsed records to Oracle database
  - Data::Standardize::Name – class for processing names
  - Data::Standardize::Address – class for processing addresses
  - Data::Standardize::Phone – class for processing phone numbers
  - Fax::DataFax::Subs – methods for Fax::DataFax
  - Fax::DataFax::OraSubs – methods for loading data to Oracle database
  - Fax::DataFax::Journal – methods for processing journal records
  - Fax::DataFax::LibSubs – methods for processing DataFax lib files
  - Fax::DataFax::DateTime – methods for calculating date and time
  - Fax::DataFax::JavaScript – methods for generating JavaScripts on-demand
  - Etc.

## Implementation - Example

```
#!/usr/local/bin/perl
# Use methods to input
use Fax::DataFax::ODM;
my $obj = new Fax::DataFax::ODM;
$obj->study_no(254);
$obj->pat_ids('1-3');
$obj->out_type('snapshot');
$obj->out_format('odm');
$obj->out_file('my_odm.xml');
$obj->batch;
```

```
#!/usr/local/bin/perl
# Use argument to input
use Fax::DataFax::ODM;
my $obj = Fax::DataFax::ODM->new;
$obj->batch(study_no=>254, pat_ids=>'1-3',
  out_type=>'snapshot', out_format=>'odm',
  out_file=>'my_odm.xml');
```

## Implementation - User Interface

- The backend ETL engine is complicated and is not user friendly.
- We developed a front-end web interface.

Enable CDISC with DataFax

Data Source:	<input checked="" type="radio"/> Local <input type="radio"/> Remote <input type="radio"/> Local time	<input type="checkbox"/> Verbose mode	<input checked="" type="checkbox"/> Format output
Select studies:	<input type="text" value="Select studies"/>	==>	<input type="text"/>
Select options:	<input type="text" value="Select options"/>	==>	<input type="text"/>
Upload:	<input type="text"/>	<input type="button" value="Browse"/>	Format: <input type="text" value="Select"/>
Actions:	Type: <input type="text" value="Select"/>	Format: <input type="text" value="Select"/>	<input type="button" value="Go"/> <input type="button" value="Reset"/> <input type="button" value="Refresh"/> <input type="button" value="Help"/>

## Implementation – Data Source

---

- **Data Sources:**
  - Local – DataFax server is on the same server as web server.
  - Remote – DataFax server is on a separated server from the web server and need to run DFgetData.
  - Real Time – Use DFexport.rpc whenever it is necessary.
- **Select Studies:**
  - You can select one or more than one studies from the list.
  - You can also use the **-s** option to specify studies, but it will be overwritten by the selection from the active study list if you have done so.

## Implementation – Options: Inputs

---

- **Set Options:**
  - Select studies
    - -s: specifies studies in Free Range Format
  - Upload additional files
    - You can upload lab or other files.

## Implementation – Options: Record Filter

---

- Set Options:
  - Record filters
    - -S: record status: primary, secondary, all, etc.
    - -I (big i): patient Ids in Free Range Format
    - -J: center numbers in FRF
    - -P: plate numbers in FRF
    - -V: visit numbers in FRF
    - -v: validation levels in FRF
    - -C: creation date range
    - -M: modification date range

## Implementation – Options: Configuration

---

- Output
  - -f: output format: SAS, XML, CSV, ODM, LAB, SDS, ADaM, etc.
  - -t: output type: transaction, snapshot, and archive
  - -O: create files by patient, center, or visit.
- Configuration: can be set in the DataFax.ini file or from the UI
  - -a: archive directory if '-t archive'
  - -b: backup/ftp location in svr:/ftp/dir:usr:pwd format
  - -d: database connection string in usr/pwd@db format for Oracle
  - -p: page directory if converting fax images to PDFs
  - -l (small L): single study with specified lib directory

## Implementation - ODM vs. DataFax

---

- ODM Study:
  - Study configuration: DFserver.cf
  - Page and variables: Dfschema, DFplates
  - Investigators: DFcenters
  - Events: DFvisit\_map, etc.
- ODM AdminData:
  - User permisisions: DFuserPerms
  - Audit trails: Journal files
- ODM ReferenceData:
  - Need to derive from other files
- ODM ClinicalData:
  - Clinical data: plate files
  - QC data: DFqc.dat

## Conclusion

---

- XML is the technology being used in many industries and being adopted in clinical trial industry as well.
- It is very important for a CRO to adopt a common and popular standard.
- CDISC provides an opportunity for us to adopt such a standard.
- XML and CDISC provide us opportunity to integrate diverse systems.

## Questions?

---

## References

---

- CDISC: <http://www.cdisc.org/>
- CDISC discussion forum:  
<http://www.cdisc.org/discussions/discussions.html>
- ODM package – software and documentation:  
[http://www.cdisc.org/discussions/odm\\_v1\\_1.html](http://www.cdisc.org/discussions/odm_v1_1.html)
- ICH: <http://www.ich.org/>
- W3C: <http://www.w3c.org/>
- HL7: <http://www.hl7.org/>
- XML: <http://www.xml.com>, <http://www.apache.org>,  
<http://alphaworks.ibm.com>, <http://www.extensibility.com>
- Office of the Assistant Secretary for Planning and Evaluation:  
<http://aspe.hhs.gov/>
- Department of Health and Human Services:  
<http://www.hhs.gov/>