

Integrating science, technology and experienced implementation

The Case for a Controlled Environment for Statistical Processing of Clinical Trial Data

Alan Hopkins, Ph.D.
October 20, 2004

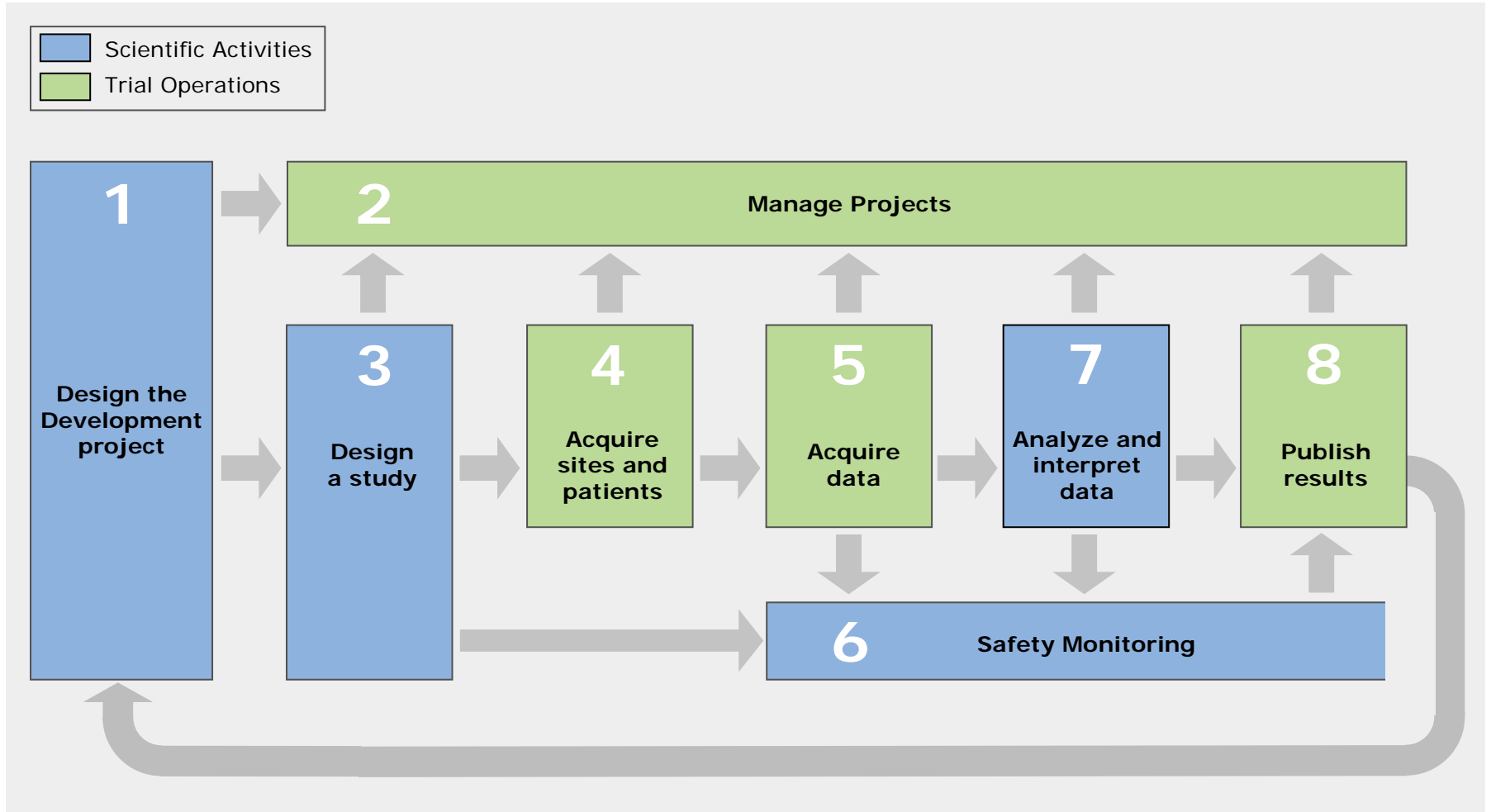
Presented at DIA-CDISC Joint Conference:
“eClinical Interchange: From Clinician to Submission”
Arlington, VA

Agenda

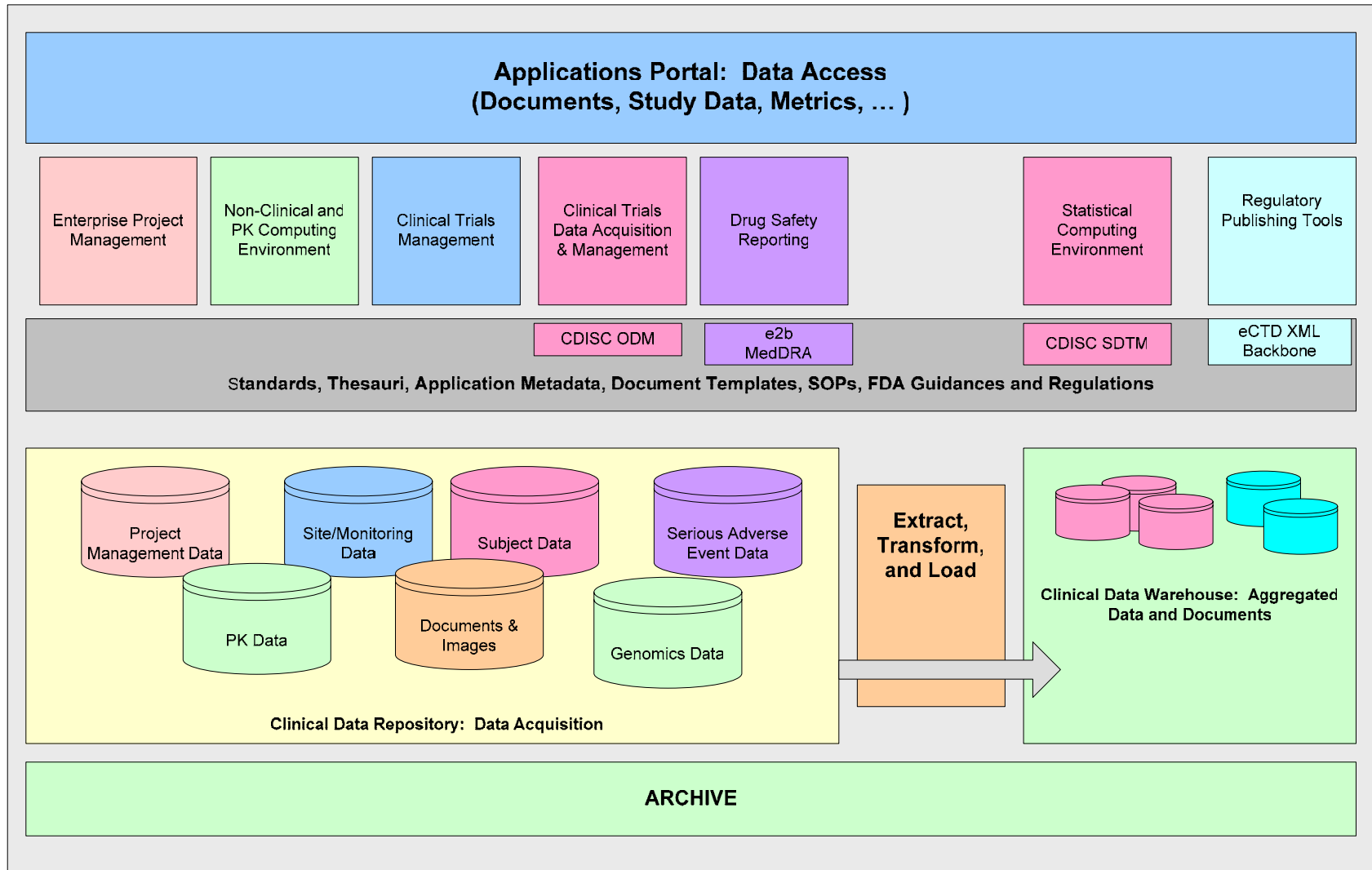
- **Context: managing the clinical data pathway**
- **Drivers for a new statistical computing process**
- **Components of a statistical computing environment**
- **The future: using the statistical analysis plan to drive automation of statistical reporting processes**
- **Summary**

▶ **Context: The Clinical Data Pathway**
Drivers for New Statistical Processes
The Statistical Computing Environment
Future Process Enhancements
Summary

The Clinical Development Process



High Level Information Systems Architecture



Integrate functional silos through the data

Effective drug development integrates science, information technology and clinical operations through effective management of the data pathway

It's all about the data – it's content, integrity, structure and documentation

▶ **Science: data content and interpretation**

- A strategic drug development plan
- Focused clinical protocols
- Statistical plans for analysis and interpretation of results

▶ **Information Technology: productivity tools**

- Data collection and storage: EDC, relational databases
- Data standards (CDISC, MedDRA, etc.)
- Analysis tools / Analysis environment
- Site management

▶ **Clinical Operations**

- Outsourcing strategy
- Program management
- Clinical site operations

Data Pathway Goal

An defined integrated process to assure an accurate, auditable data pathway exists for clinical information that flows into electronic regulatory submissions.

Effective management of the data pathway must be a core competence of all development organizations

Drivers for the business case to manage the entire data pathway as one process

- Data is the most important asset of a clinical development organization
- A holistic process approach to quality facilitates assessment of risks
- Integrate functional silos to enhance teamwork and support best practices
- Provides a basis for training
- Sets expectations for the entire organization
- The statistical integrity of data analyses and the reporting of results should be a top priority.

We will focus on the processes related to statistical analysis.

Context: The Clinical Data Pathway
▶ Drivers for New Statistical Processes
The Statistical Computing Environment
Future Process Enhancements
Summary

Multiple drivers for creation of statistical environments

- FDA Regulations/Guidelines
 - ICH E9: Statistical Principles for Clinical Trials
 - 21 CFR Part 11
 - Part 11, Electronic Records; Electronic Signatures: Scope and Application
 - “Computerized Systems Used in Clinical Trials” (April 1999)
 - eCTD
- Data Standards
 - CDISC: ODM, SDTM, ADaM, HL7 Protocol Representation
- Statistical analysis plans (ICH E3, HL7 subcommittee)
- Increasing Complexity
 - Technologies: XML, SDTM, CMS, SAS, S-PLUS
- Outsourcing and data sharing (CROs, Labs, DMCs, Partners, FDA, ...)
- Cost of validation – what can be done to minimize cost
- Good business (and science) practices

Good Statistical Practice

Guidance for Industry

E9 Statistical Principles for Clinical Trials

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
September 1998
ICH

Accuracy and reliability

- Minimize bias
- Maximize precision

Validity and Integrity

- Prespecified analysis – SAP
- Integrity of data and statistical software

“The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software ... used both for data management ... and for processing the data statistically.”

Computerized Systems Used In Clinical Trials

Guidance for Industry

COMPUTERIZED SYSTEMS USED IN CLINICAL TRIALS

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Biologic Evaluation and Research (CBER)
Center for Drug Evaluation and Research (CDER)
Center for Devices and Radiological Health (CDRH)
Center for Food Safety and Nutrition (CFSAN)
Center for Veterinary Medicine (CVM)
Office of Regulatory Affairs (ORA)
April 1999

Computerized System

means, for the purpose of this guidance, computer hardware, software, and associated documents (e.g., user manual) that *create, modify, maintain, archive, retrieve, or transmit* in digital form information related to the conduct of a clinical study.

Management of Electronic Records and Signatures

Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application

*Division of Drug Information, HFD-240
Center for Drug Evaluation and Research (CDER)
(Tel) 301-827-4373
<http://www.fda.gov/cder/guidance/index.htm>
or
Office of Communication, Training and
Manufacturers Assistance, HPM-40
Center for Biologics Evaluation and Research (CBER)
<http://www.fda.gov/cber/guidelines.htm>
Phone: the Voice Information System at 800-835-4709 or 301-827-1800
or
Communications Staff (HFV-12),
Center for Veterinary Medicine (CVM)
(Tel) 301-594-1755
<http://www.fda.gov/cvm/guidance/guidance.html>
or
Division of Small Manufacturers Assistance (HFZ-226)
<http://www.fda.gov/cdrh/ggpm.htm>
Manufacturers Assistance Phone Number: 800.638.2041 or 301.443.6597
Internet Staff Phone: 301.827.3993
or
Center for Food Safety and Applied Nutrition (CFSAN)
<http://www.efsa.fda.gov/~dms/guidance.html>.*

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Center for Food Safety and Applied Nutrition (CFSAN)
Center for Veterinary Medicine (CVM)
Office of Regulatory Affairs (ORA)

August 2003
Pharmaceutical CGMPs

Applies to records in electronic form that are *created, modified, maintained, archived, retrieved, or transmitted*, under any records requirements set forth in agency regulations

“...we do not intend to take enforcement action to enforce compliance with the validation, audit trail, record retention, and record copying requirements of part 11 as explained in this guidance.”

There is no FDA guidance specifically directed toward e-record keeping for statistical analyses

What is an e-record for statistical analyses?

- Is each execution of the program an e-record?
- Is the final analysis in the submission to FDA the e-record?
- Is the statisticians report the only e-record?
- Are tests of the assumptions of ANOVA e-records?
- Are preliminary analyses in model building e-records?
- Are interim analyses e-records?
- Reference: J. McCormack (2002) Statistical Software Validation: Regulatory Perspective, Society for Clinical Trials Annual Meeting.

Simple solution: Manage all e-records in an environment that tracks these objects seamlessly as part of the ordinary work environment

Part 11 Presupposes Operational Maturity

- Have a goal and plan for everything you do
- Actively control how tasks get done
- Control access and changes to records and documents
- Maintain audit trails for records and documents
- Exercise commercial discipline in the development and use of software
- Exercise appropriate security for systems and data
- Make sure everyone is trained and qualified for their tasks

Study Data Tabulation Model: 200 Pages of Explanation

Study Data Tabulation Model

Prepared by the

CDISC Submission Data Standards Team

Principal Editor: Wayne Kutick
Principal Contributors: Fred Wood, Diane Wolf, Tom Quinter, Julie Evans, CDISC SDS Team

Notes to Readers

This is the released Version 1.0 of the Study Data Tabulation Model Document, previously posted for comment by the CDISC Submission Data Standards team. This document, which supersedes all prior versions, reflects changes from two comment periods: an initial comment period in March/April 2004 through CDISC and the HL7 Regulated Clinical Research Information Management Technical Committee, and a second review period from May 27 to June 10, 2004 through CDISC.

Revision History

Date	Version	Summary of Changes	Primary Author
20040625	Version 1.0	Released version reflecting all changes identified during comment periods.	Kutick, Wood, Evans, Wolf, Quinter

CDISC © 2004. All rights reserved.
Final

Page 1
June 25, 2004

CDISC SDTM Implementation Guide (SDS Version 3.1)



Study Data Tabulation Model Implementation Guide: Human Clinical Trials

Prepared by the
CDISC Submission Data Standards Team

Notes to Readers

- This is the approved implementation guide for Version 1 of the CDISC Study Data Tabulation Model.
- This Implementation Guide comprises version 3.1 of the Submission Data Standards.

Revision History

Date	Version	Summary of Changes
2004-05-25	Document Version 1.0	Released version reflecting all changes identified during comment periods.
2004-07-14	Document Version 1.01	Corrects minor typos and errors in sections 4.1.4.2, 5.1.1.0.7, 5.1.2.0, 6.3.1, 6.3.3, 6.3.4, 6.3.6.

CDISC © 2004. All rights reserved.
Final

Page 1
July 14, 2004

Findings Datasets – Labs example

Highly normalized datasets need to be reconfigured for statistical analysis.

	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBSCAT	LBORRES	LBORRESU	LBORNRHI	LBORNRLO
Row 1	ABC	LB	ABC-001-001	1	ALB	Albumin	CHEMISTRY		34	g/L	50	35
Row 2	ABC	LB	ABC-001-001	2	ALKP	Alkaline Phosphatase	CHEMISTRY		398	IU/L	160	40
Row 3	ABC	LB	ABC-001-001	3	ALKP	Alkaline Phosphatase	CHEMISTRY		350	IU/L	160	40
Row 4	ABC	LB	ABC-001-001	4	ALKP	Alkaline Phosphatase	CHEMISTRY			IU/L		
Row 5	ABC	LB	ABC-001-001	5	WBC	WBC Count	HEMATOLOGY		5.9	10*9/L	11	4
Row 6	ABC	LB	ABC-001-001	6	LYMPH	Lymphocytes	HEMATOLOGY	DIFFERENTIAL	6.7	%	40	25
Row 7	ABC	LB	ABC-001-001	7	NEUT	Neutrophils	HEMATOLOGY	DIFFERENTIAL	5.1	10*9/L	8	2
Row 8	ABC	LB	ABC-001-001	8	PH	pH	URINALYSIS		7.5		9.0	5.0
Row 9	ABC	LB	ABC-001-001	9	ALB	Albumin	CHEMISTRY					
Row 10	ABC	LB	ABC-001-001	10	CHOL	Cholesterol	CHEMISTRY		229	mg/dL	199*	0*
Row 11	ABC	LB	ABC-001-001	11	WBC	WBC Count	HEMATOLOGY		5.9	10*9/L	11	4

Source: CDISC SDTM

Statistical Analysis Dataset Model



Statistical Analysis Dataset Model: General Considerations Version 0.7

Prepared by the
CDISC Analysis Dataset Modeling Team
(ADaM)

Notes to Readers

- This Model supersedes all previous ADaM models
- Additional models for specific statistical methods will be developed using the concepts and standards presented in this document

Revision History

Date	Version	Description
9/15/05	Draft 0.7	Draft Version for public review

General description of statistical analysis datasets used introduces

Value-level metadata

Variable level metadata

Analysis level metadata

“The general concepts, descriptions and definitions in this document will be used to build subsequent statistical model documents.”

Independent data monitoring committees add value and complexity

Guidance for Clinical Trial Sponsors

On the Establishment and Operation of Clinical Trial Data Monitoring Committees

DRAFT GUIDANCE

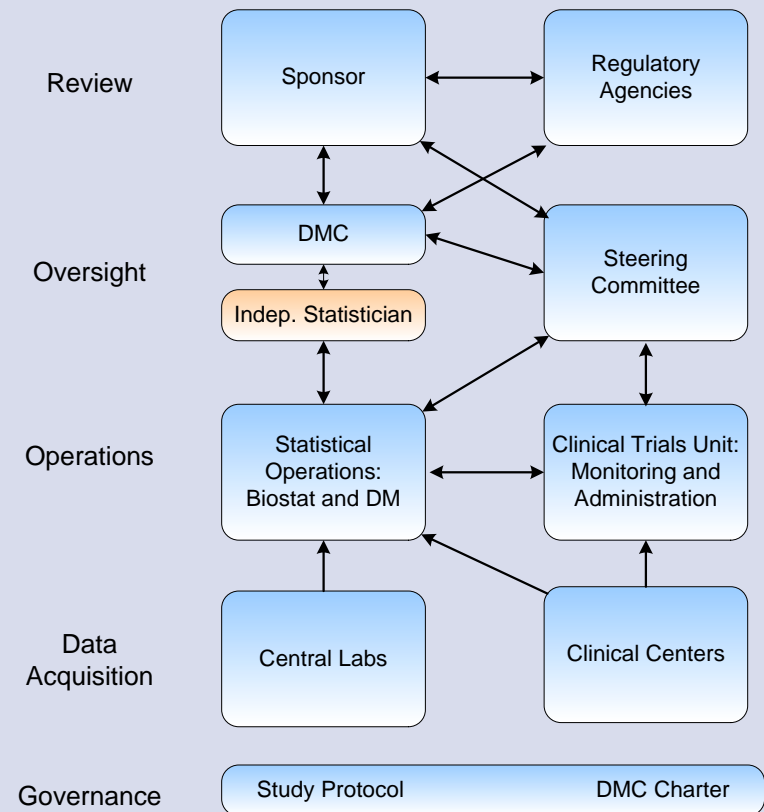
This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted by the date provided in the *Federal Register* notice announcing the availability of the draft guidance. Submit comments to Docket Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that published in the *Federal Register*.

For questions on the content of this draft document contact Mary Feulkes (CBER), 301-827-3034, Robert Temple (CDER), 301-594-6758, or Joanne Less (CDRH), 301-594-1190.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Biologics Evaluation and Research (CBER)
Center for Drug Evaluation and Research (CDER)
Center for Devices and Radiological Health (CDRH)
November 2001

Typical Phase 3 Trial Organization



The electronic common technical document

Guidance for Industry

Providing Regulatory Submissions in Electronic Format — Human Pharmaceutical Product Applications and Related Submissions

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.










Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (CDER) Randy Levin 301-594-5411, or (CBER) Robert Yetter at 301-827-0373.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

August 2003
Electronic Submissions

\\CDER\CDER\GUIDE\5640.dft.doc
08/18/03

 { module }	Replace with module name, e.g., m5
 datasets	
 {study}	Replace with study identifier, e.g., 123-070
 analyses	Contains analysis datasets, annotated CRF, data definition
 programs	Contains program files
 ecgs	Contains annotated ECG waveform datasets
 listings	Contains data listing datasets, annotated CRF, data definition
 profiles	Contains subject profiles
 tabulations	Contains data tabulation datasets, annotated CRF, data definition

FDA wants multiple types of data files, documentation, and programs – the whole statistical environment

Context: The Clinical Data Pathway
Drivers for New Statistical Processes

▶ **The Statistical Computing Environment**

Future Process Enhancements

Summary

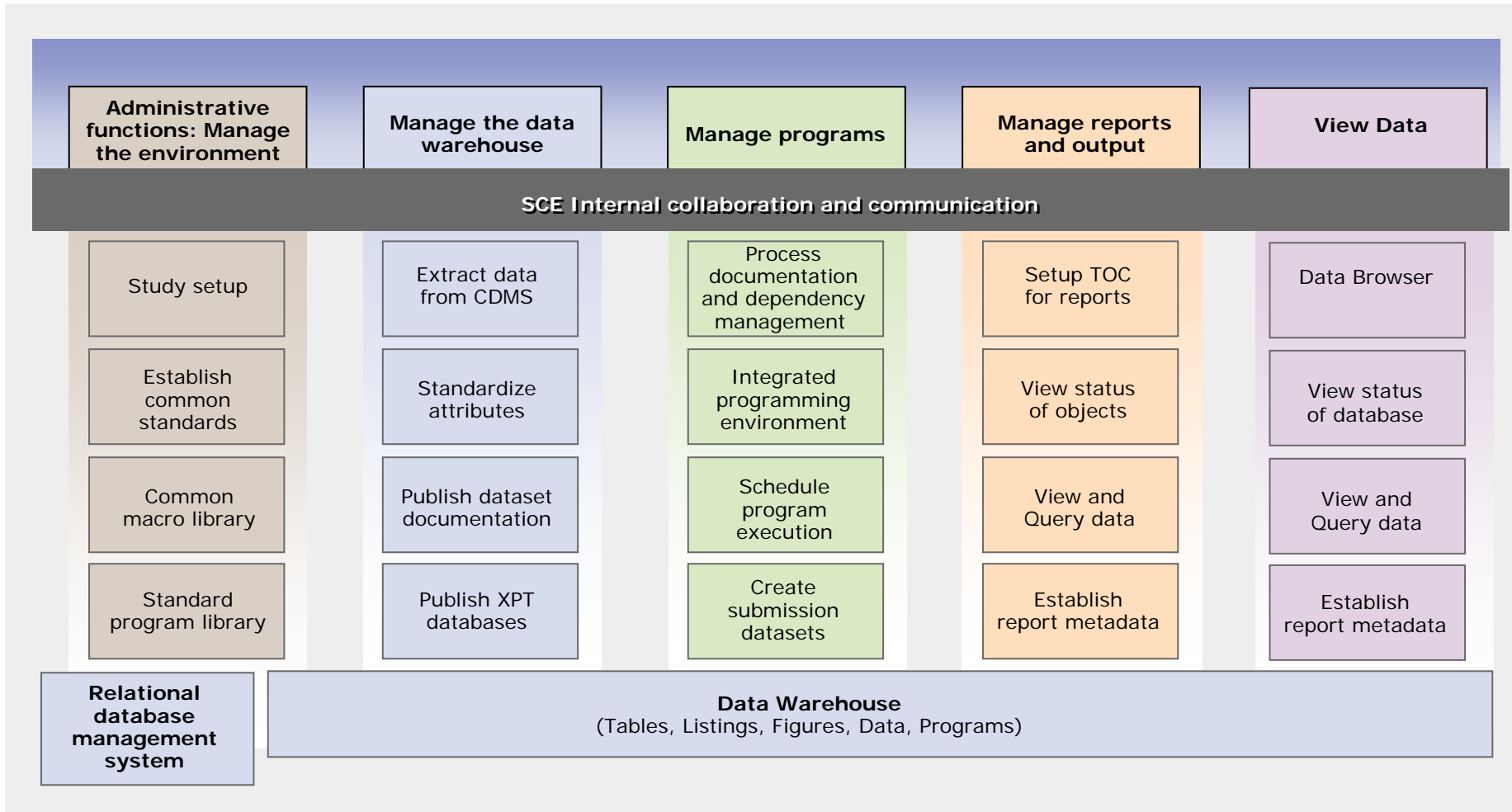
What is a statistical computing environment?

A Statistical Computing Environment (SCE) provides a foundation for documenting rigor in the analysis and reporting of clinical trial results while increasing productivity and quality.

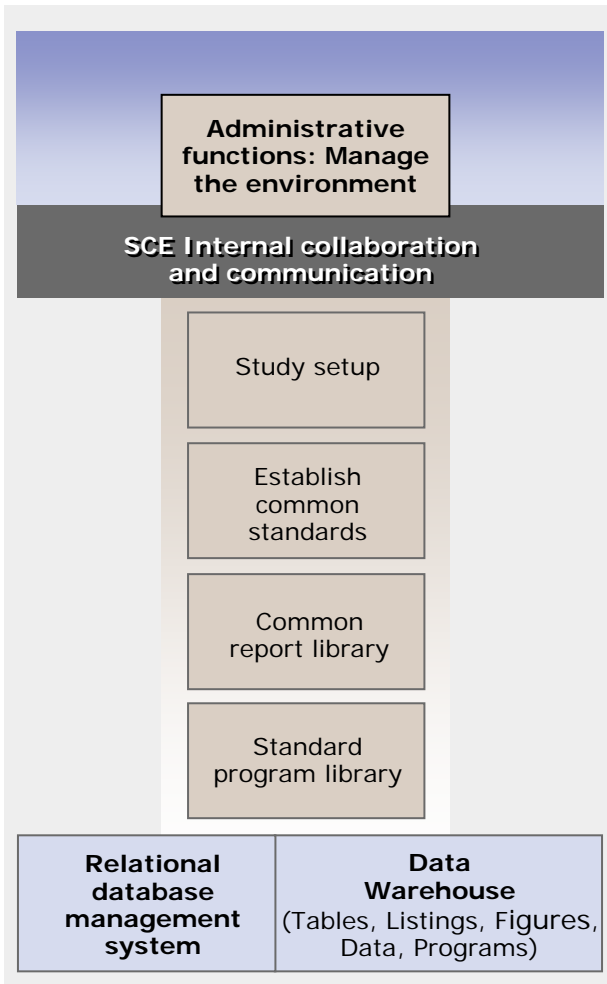
Characteristics of a Statistical Computing Environment

- Tools are targeted directly to the deliverables (clinical study reports, analysis files, eCTD, etc) necessary for regulatory submissions.
- It organizes the various activities, breaking down the whole process into smaller, simpler tasks.
- The whole process becomes transparent. It is easier to train new workers and to track progress on large projects.
- A shared platform usable by both programmers and statisticians facilitates communication and productivity of all concerned.
- Documentation is created by a structured approach.
- Creates a process that makes Part 11 compliance a by-product of work not an object of work.

The Statistical Computing Environment: Conceptual Components

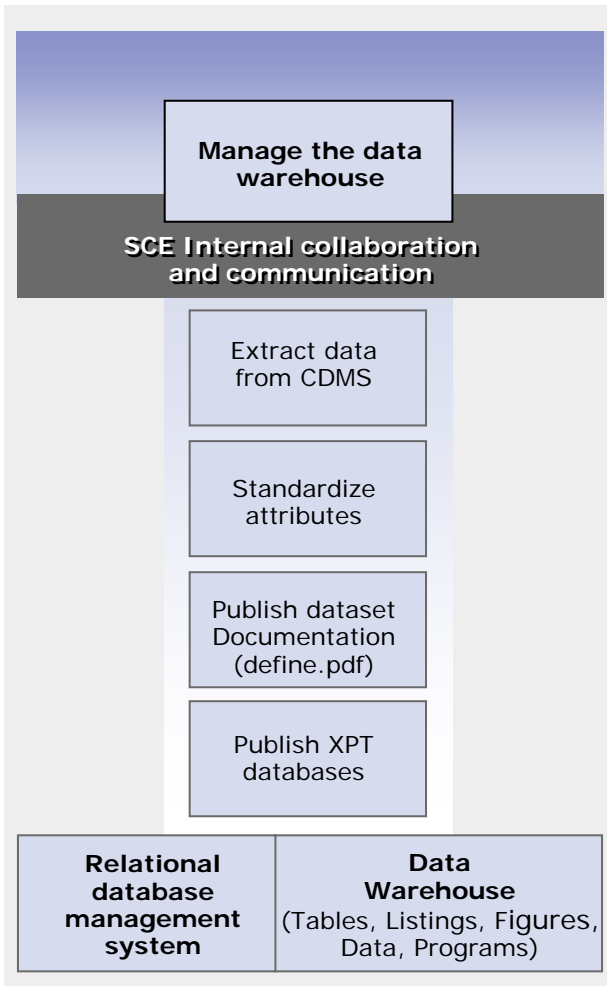


Manage the Computing Environment



- Add users, define user roles
- Define directory structures and permissions
- Store templates
- Management of standardized reporting software e.g. SAS and S-Plus programs
- Control business rules
- View/report project status based on metadata
 - Catalog of programs and their status
- Export or archive data, programs, results

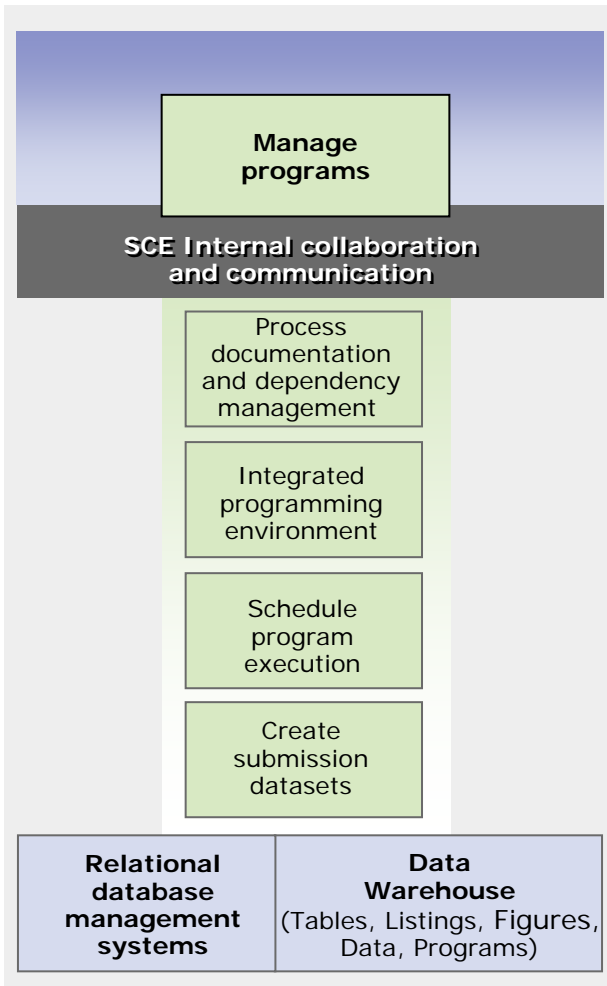
Manage the data warehouse



A repository for documents, data, programs, and output

- Support for import of SDTM data from clinical data management or vendor
- Provide data security and audit trails for the data warehouse
- Store data, logs, programs and output as BLOBs in a RDBMS. Handle these components as related objects.
- Store statistical analysis files

Manage analysis programs

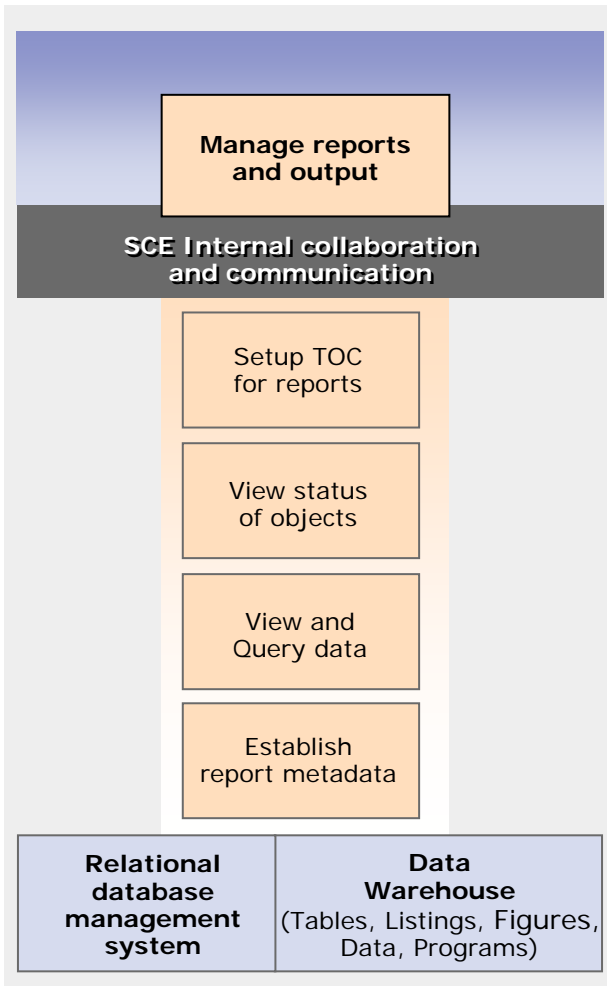


- Code management: Check-in / check-out
- Audit trails
- Batch processing
- Dependency management
- Extensible: supports multiple tools
- Flexible: allows program execution outside the environment for early phases of development

Common code library provides building blocks for efficient reporting

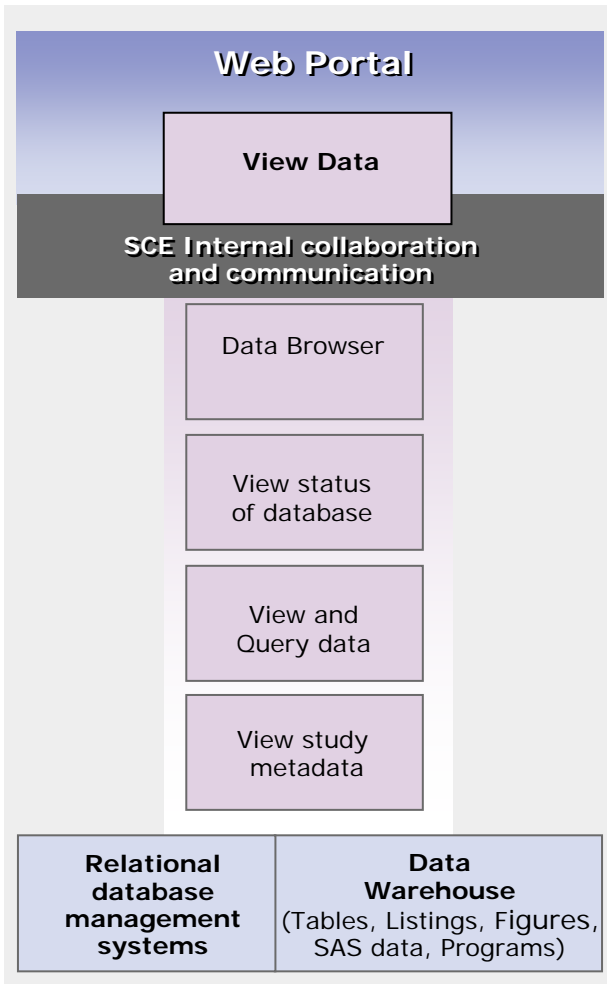
- Simplifies program development – saves time
 - Code is easier to maintain
 - Validated building blocks save time
- Software validation costs are high for statistical tables and graphs because study reports usually have a large number of custom reports.
 - Science dictates that most clinical trials are not clones of each other and so trials require custom programming for data management and analysis.
 - Systems should be deployed which minimize the time and cost of validation
 - Most system validation documentation should be generated automatically.

Manage reports and output



- Manage output through a table of contents associated with the final report
- Life cycle management for tables and graphs
 - eg. Draft --> validated --> final
- Export/Transport: data, programs
- Archival

Tools to View Data



SDTM data is not readily usable in statistical software nor for viewing in context of other data

- A tool is necessary to view SDTM data
- Data stored in a relational database
- View study metadata
- Create domain listings
- View patient profiles
- Create graphs
- Data restructuring and export

SCE Advantages

- Productivity enhancement
- Compliance with Part 11
- Facilitates Training
- Manage objects
 - Analysis files
 - Programs
 - Reports
- Process documentation
- Audit trail
- File security
- Version control
- Study archival

Context: The Clinical Data Pathway
Drivers for New Statistical Processes
The Statistical Computing Environment

▶ **Future Process Enhancements**
Summary

HL7 Protocol Representation Standards

- **Goal: Publication of a standard, machine-readable model for protocol representation that will facilitate interchange of this data among systems and stakeholders**
- **Reusability of information through XML documents, e.g.**
 - **Study design**
 - **Schedule of events**
- **Standard for statistical analysis plans**

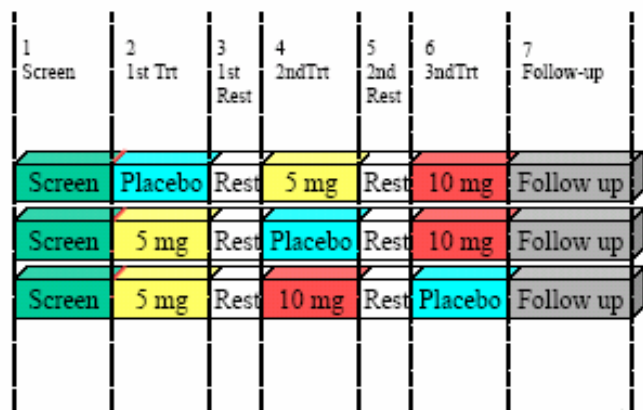
Data-based protocol representation enables automation tools for down-stream statistical processes

Consider the possibilities:

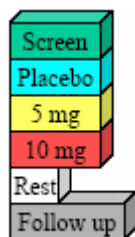
- An electronic statistical analysis plan
- Description of the appropriate analysis files
- Definition of displays of study information in final reports
- Automatic generation of programs for study reporting
- Final study report: integration of protocol elements and statistical reports

The Trial Design Model: Treatment Arms, Elements, and Epochs

Treatment Arms and Epochs:



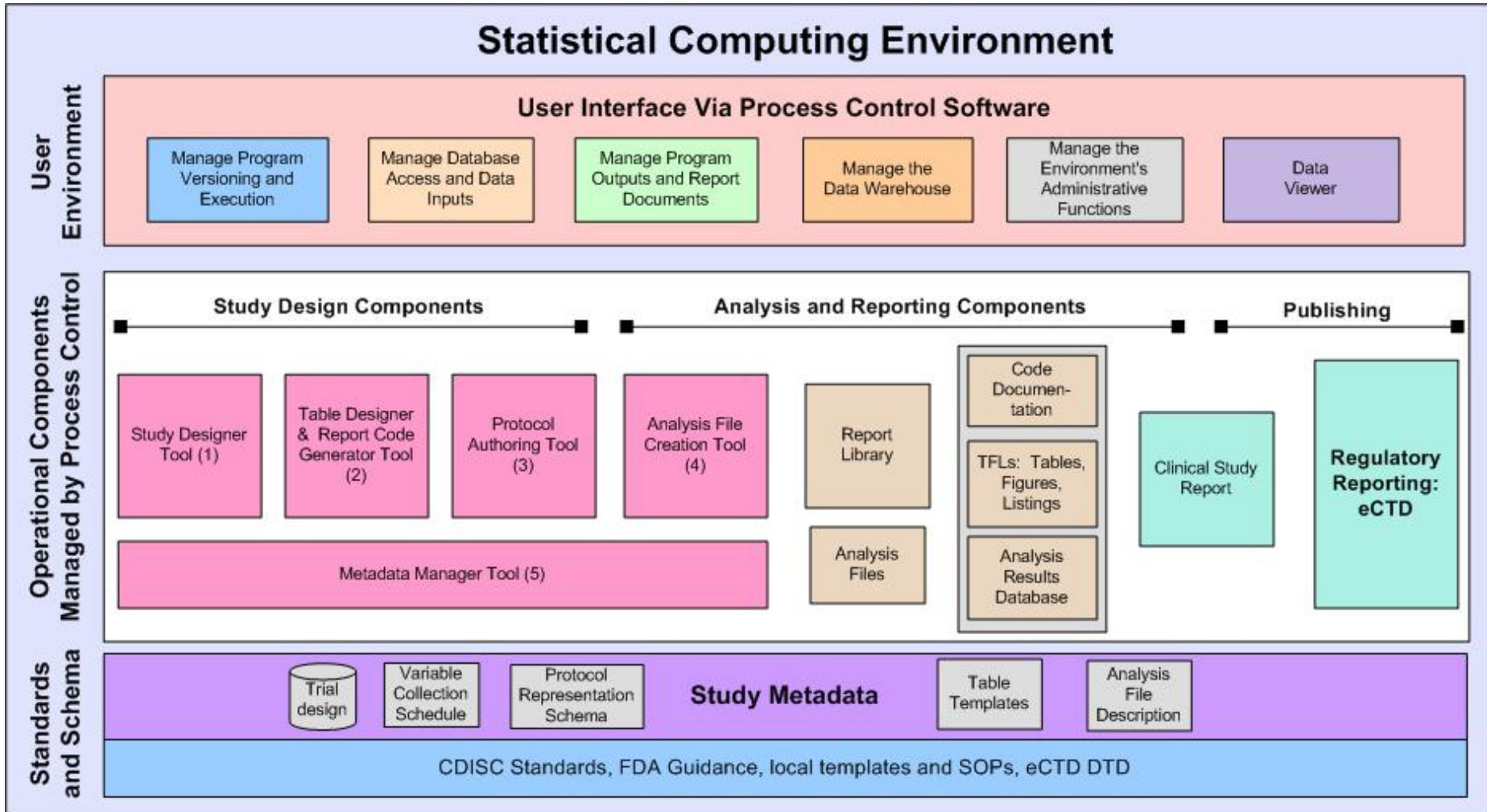
Elements:



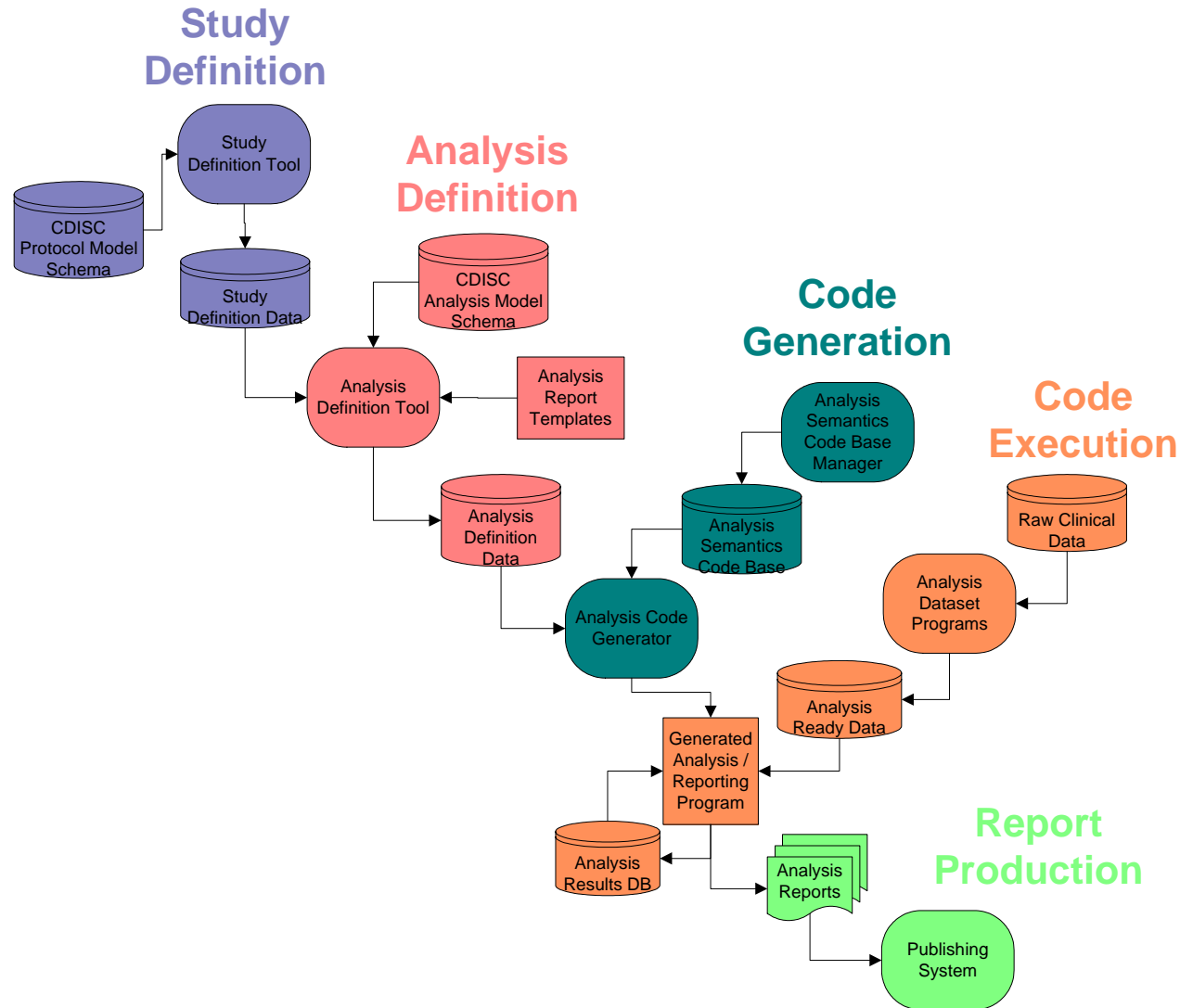
ARM	TAETORD	ELEMENT	TESTRL	TEDUR	TABRANCH
P-5-10	1	Screen	Informed consent	14D	Randomized to Placebo/5 mg/10 mg
P-5-10	2	Placebo	First dose of placebo	14D	
P-5-10	3	Rest	48 hrs after last dose drug	7D	
P-5-10	4	5 mg	First dose of 5 mg drug	14D	
P-5-10	5	Rest	48 hrs after last dose drug	7D	
P-5-10	6	10 mg	First dose of 10 mg drug	14D	
P-5-10	7	Follow-up	48 hrs after last dose drug	21D	
5-P-10	1	Screen	Informed consent	14D	Randomized to 5 mg/Placebo/10 mg
5-P-10	2	5 mg	First dose of 5 mg drug	14D	
5-P-10	3	Rest	48 hrs after last dose drug	7D	
5-P-10	4	Placebo	First dose of placebo	14D	
5-P-10	5	Rest	48 hrs after last dose drug	7D	
5-P-10	6	10 mg	First dose of 10 mg drug	14D	
5-P-10	7	Follow-up	48 hrs after last dose drug	21D	
5-10-P	1	Screen	Informed consent	14D	Randomized to 5 mg/10 mg/Placebo
5-10-P	2	5 mg	First dose of 5 mg drug	14D	
5-10-P	3	Rest	48 hrs after last dose drug	7D	
5-10-P	4	10 mg	First dose of 10 mg drug	14D	
5-10-P	5	Rest	48 hrs after last dose drug	7D	
5-10-P	6	Placebo	First dose of placebo	14D	
5-10-P	7	Follow-up	48 hrs after last dose drug	21D	

Source: CDISC SDTM

Process Automation Suite Components within a Clinical Trials Statistical Computing Environment



Flow of logic and components of an well-defined process for preparing statistical reports



Context: The Clinical Data Pathway
Drivers for New Statistical Processes
The Statistical Computing Environment
Future Process Enhancements

▶ **Summary**

Summary: The clinical data pathway

Effective drug development integrates science, technology and clinical operations through effective management of the data pathway

Effective management of the data pathway must be a core competence of all development organizations

Manage information flow to assure an accurate, auditable data pathway exists for clinical information that flows into electronic regulatory submissions to support data integrity.

Summary: Statistical Computing Process

A structured statistical computing environment provides a foundation for documenting rigor in the analysis and reporting of clinical trial results while increasing productivity and quality.

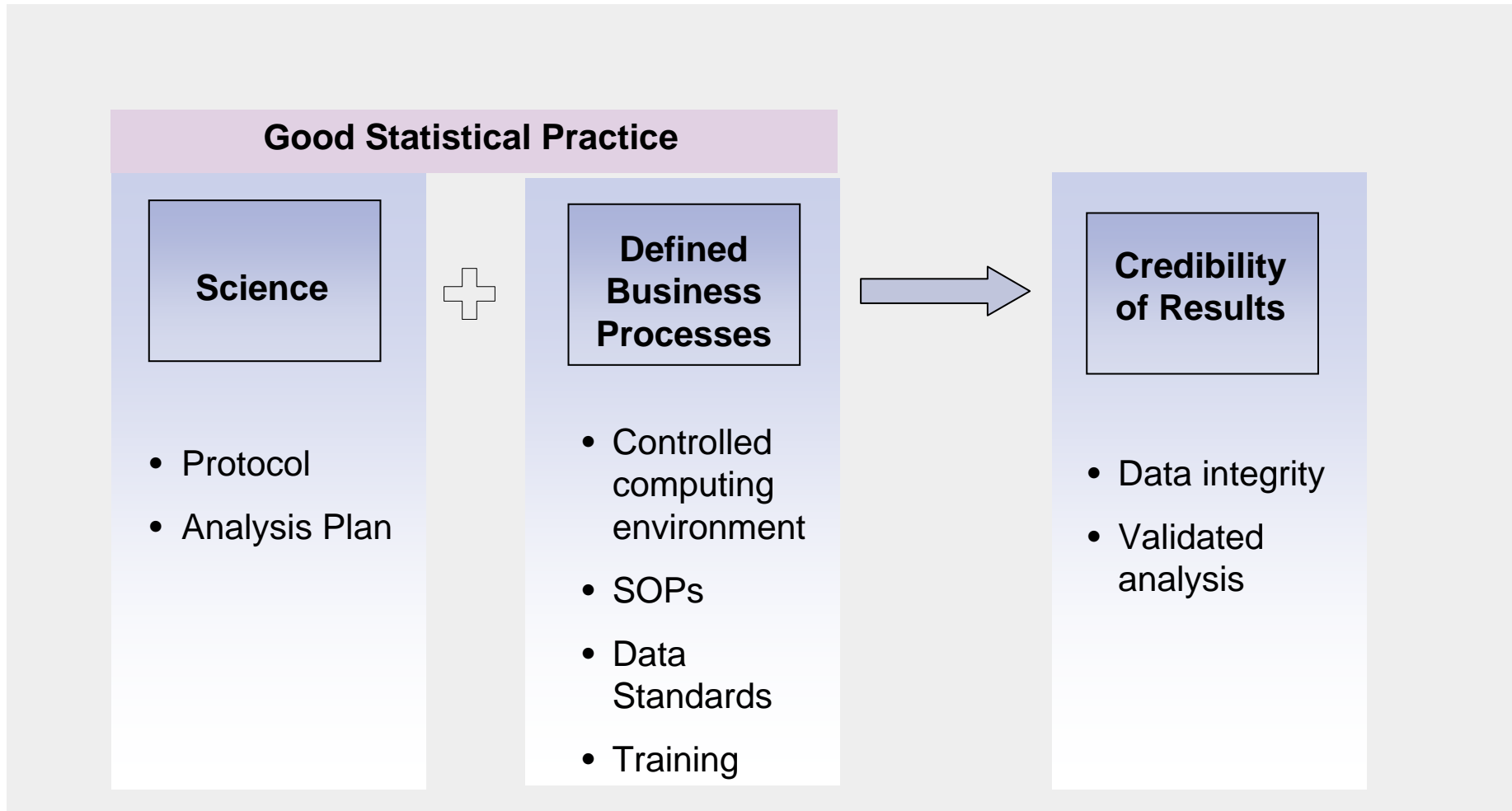
- Build a structured environment that supports multiple tools
 - Governed by SOPs, e.g. validation
 - Part 11 compliant: audit trails, code management, data security
- A lot of standards and guidances to keep your eyes on
- Manage the complexity
- Need a specification language for statistical tables
- An SCE gives a foundation for documenting rigor in the analysis and reporting process of clinical trial results while increasing productivity and quality.

Summary: Tools for the new processes

New data standards provide an opportunity to structure traditional processes for better communications within drug development teams, regulatory agencies, and vendors who will create the new tools needed by all involved.

- Well-defined processes can communicate through XML
- New tools are needed
 - Protocol authoring
 - SAP authoring
 - Specification of statistical analysis files
 - Definition of statistical tables

SCE and other process enhancements support *Good Statistical Practice*



Contact Information

Alan Hopkins

ahopkins @ pharmastat.com

925-631-9400