

Integrating science, technology and experienced implementation

# Statistical Computing Environments: What It Is, Why It's Hot

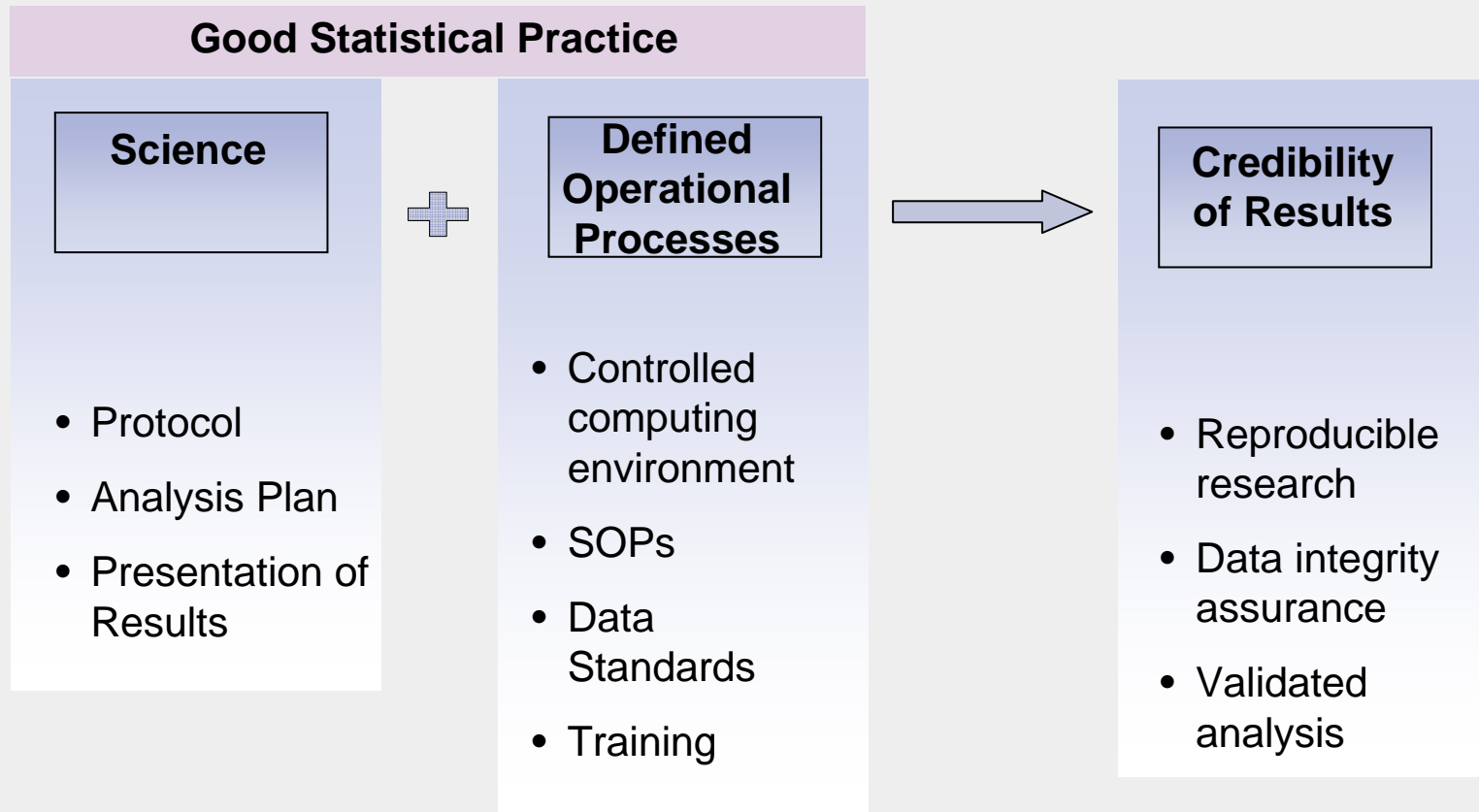
**Alan Hopkins, Ph.D.**  
**June 27, 2005**

**Presented at DIA Annual Meeting 2005**

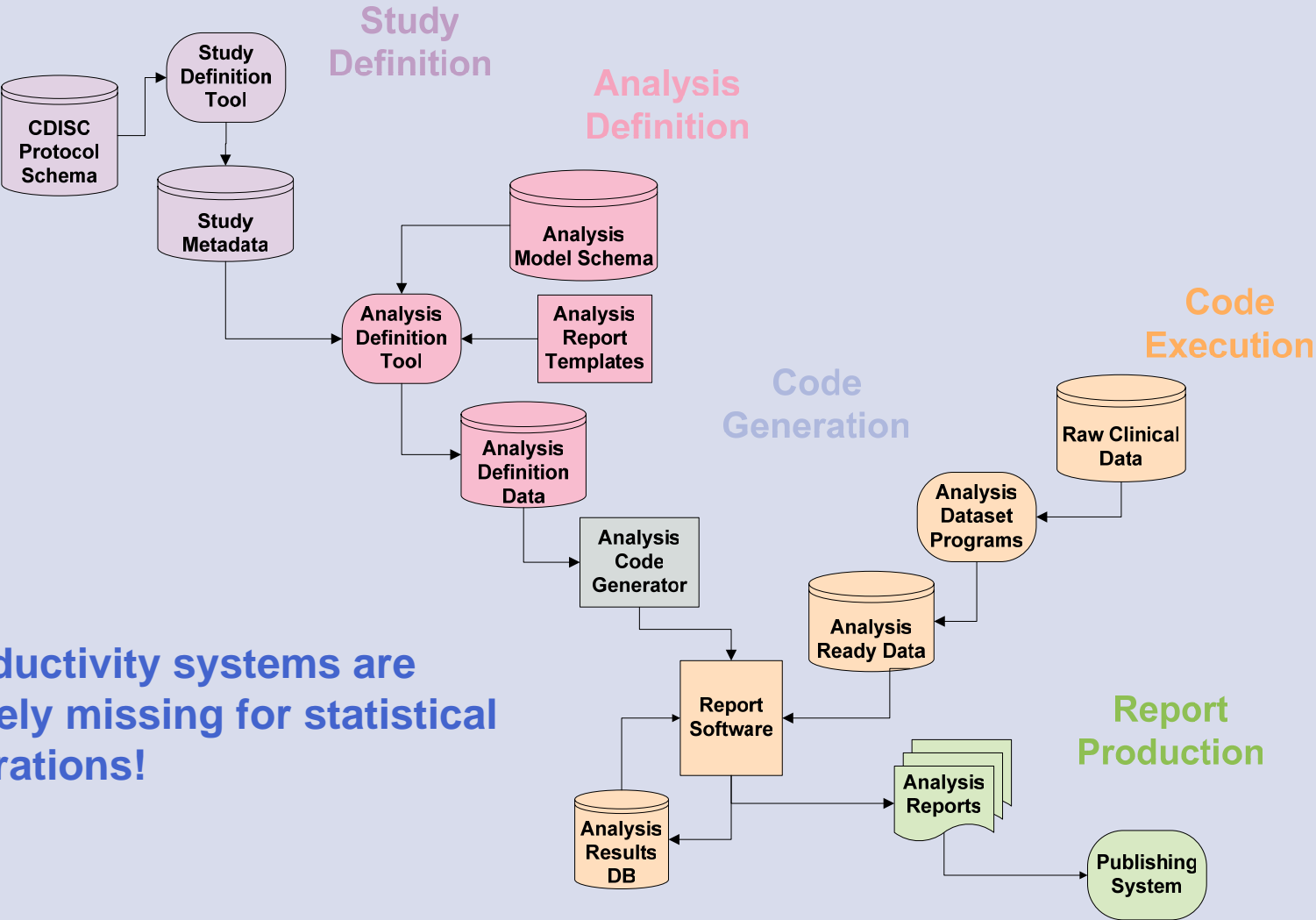
**Session: “Tools Sponsors Need to Implement Industry-wide Data Standards ”**

**Washington, DC**

# A Structured Statistical Computing Environment (SCE) and other process enhancements support *Good Statistical Practice*



# Statistical process components for preparing clinical trial reports



Productivity systems are largely missing for statistical operations!

# Multiple drivers for creation of statistical environments

***Because the clinical trials data have broad public health significance, the data are expected to be of the highest quality and integrity***

## ▶ **FDA Regulations/Guidelines**

- ICH E9: Statistical Principles for Clinical Trials
- 21 CFR Part 11 & Scope and Application Guidance
- “Computerized Systems Used in Clinical Trials” (April 1999; new draft Sept 2004)
- eCTD specification

## ▶ **Data Standards**

- CDISC: ODM, SDTM, ADaM, HL7 Protocol Representation

## ▶ **Statistical analysis plans (ICH E3, CDISC/HL7 subcommittee)**

## ▶ **Increasing Complexity**

- **Technologies:** XML, CMS, SAS, S-PLUS
- **Outsourcing and data sharing** (CROs, Labs, DMCs, Partners, FDA, ...)

## ▶ **Cost of validation** – what can be done to minimize cost?

## ▶ **Other good business (and science) practices** – Literate Statistical Practice

# 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/gmpmain.html>  
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.”

## Part 11 presupposes operational maturity

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

# 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

# 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



## 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.

# Statistical analysis data files – The key thing is simplicity documentation



## 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

The data are analysis ready

Documentation of statistical analysis datasets:

- Value-level metadata
- Variable level metadata
- Analysis level metadata

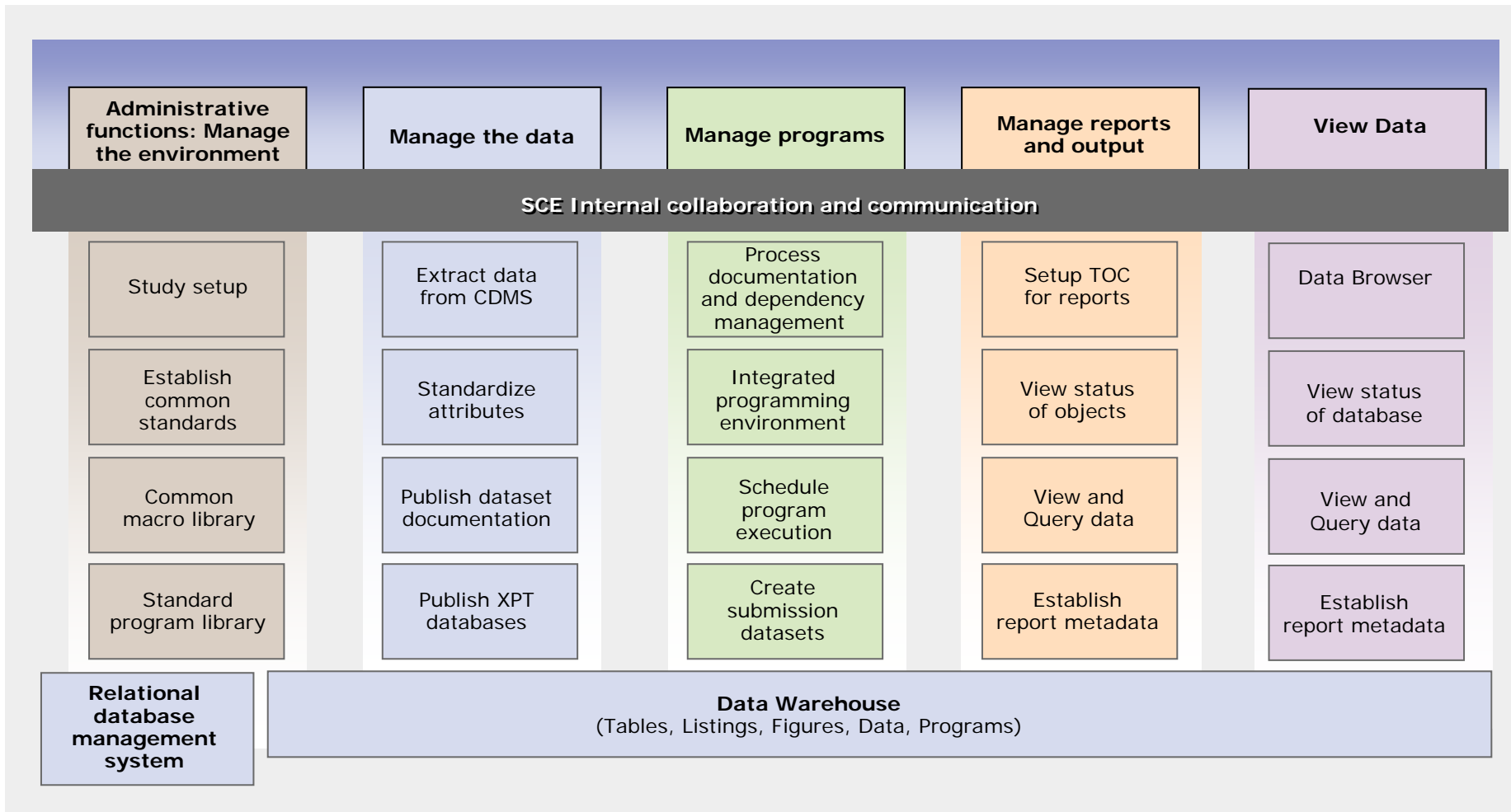
“The general concepts, descriptions and definitions in this document will be used to build subsequent statistical model documents.”

# Characteristics of a Statistical Computing Environment

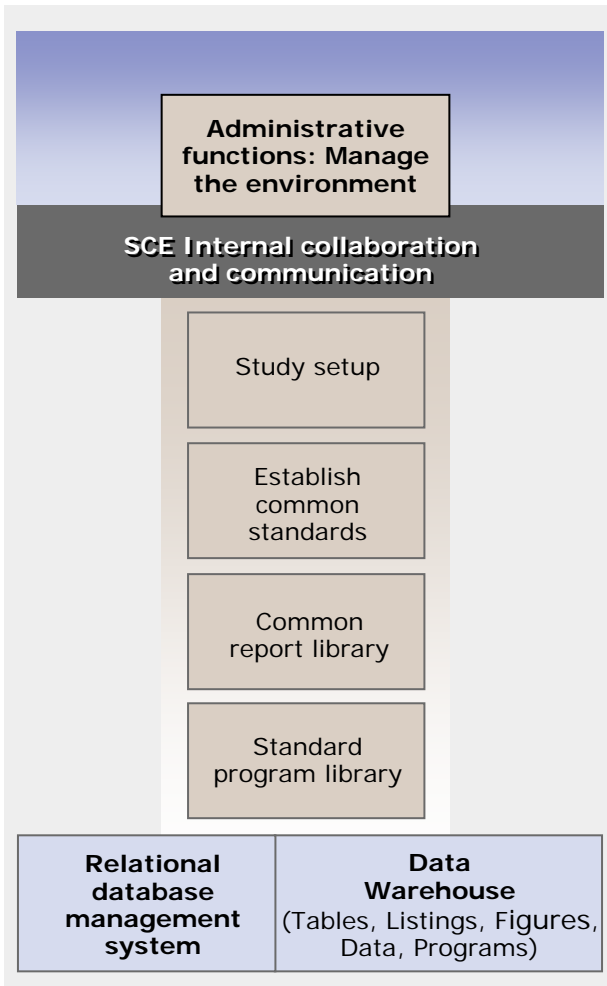
***A Statistical Computing Environment (SCE) provides a foundation for documenting rigor in the analysis and reporting of clinical trial results while increasing productivity and quality.***

- 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

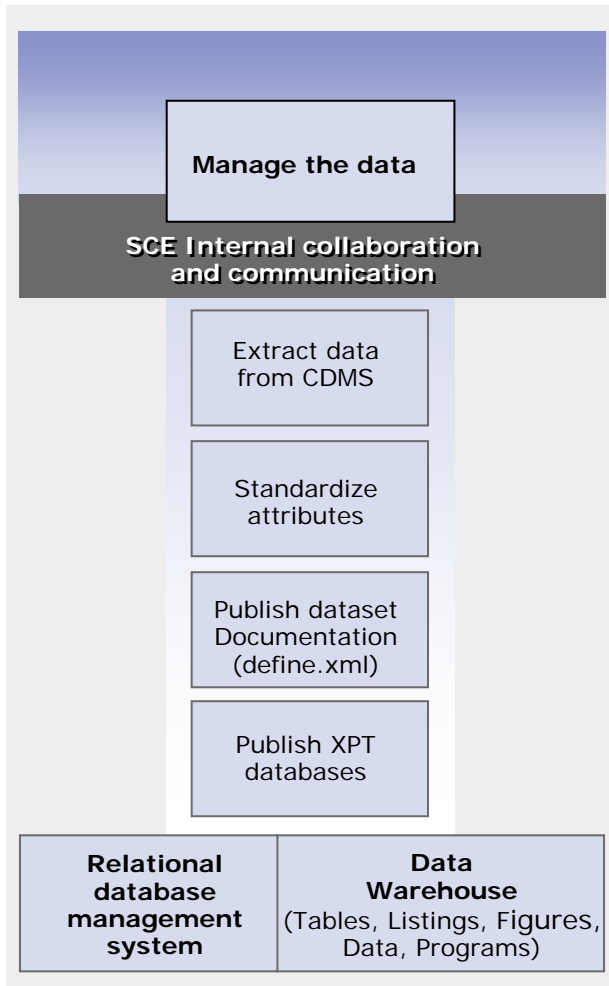


# Manage the Computing Environment



- 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

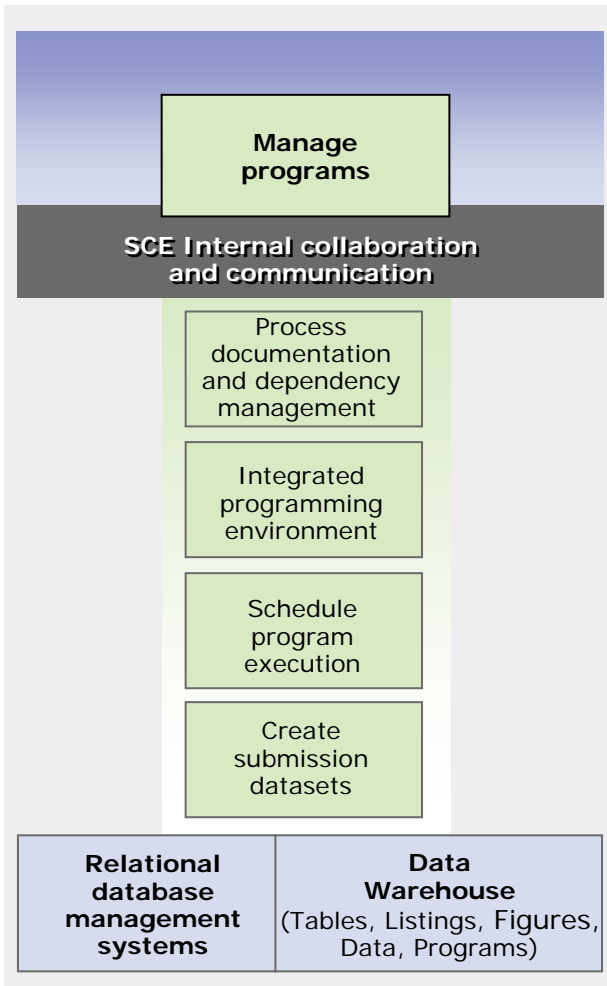
# Manage the data



A repository for documents, data, programs, and output

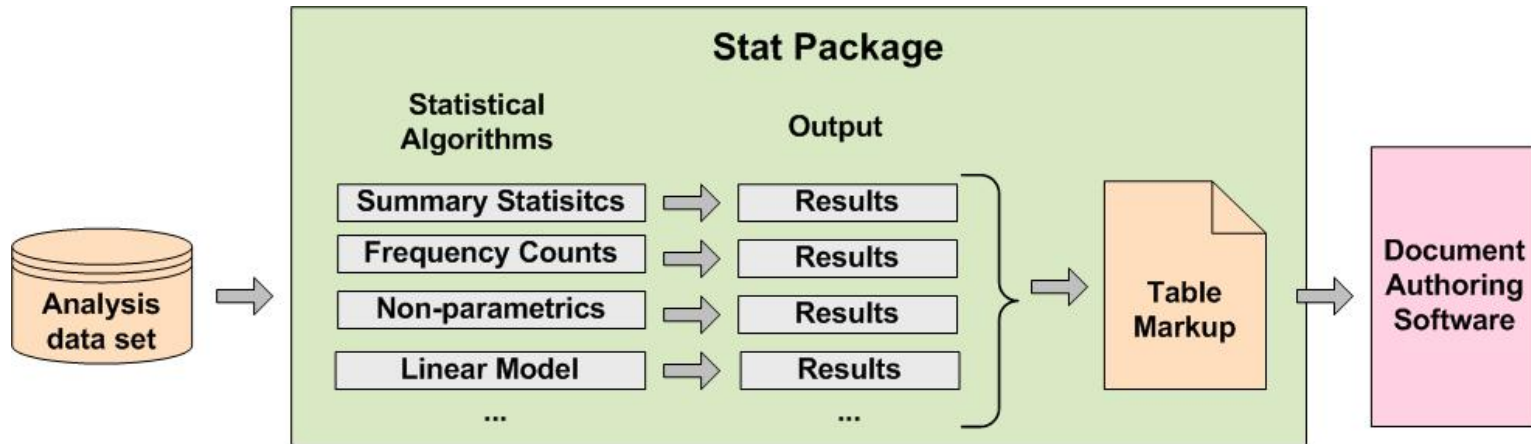
- Support for import of SDTM data from clinical data management or vendor
- Provide data security and audit trails for the data warehouse
- Store data, logs, programs and output as BLOBs in a RDBMS. Handle these components as related objects.
- Store statistical analysis files

# 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

# Statistical Programming is often a bottleneck



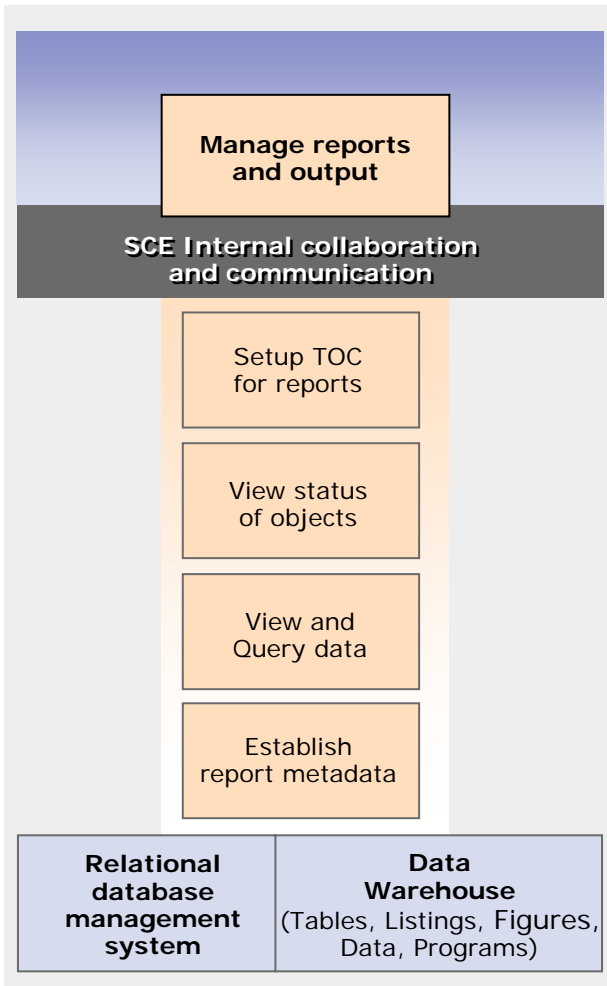
## Today

- Substantial programming often required
- Common code library provides building blocks for statistical tables
  - Code is easier to maintain
  - Validated building blocks save time
  - Facilitates training

## What's Needed

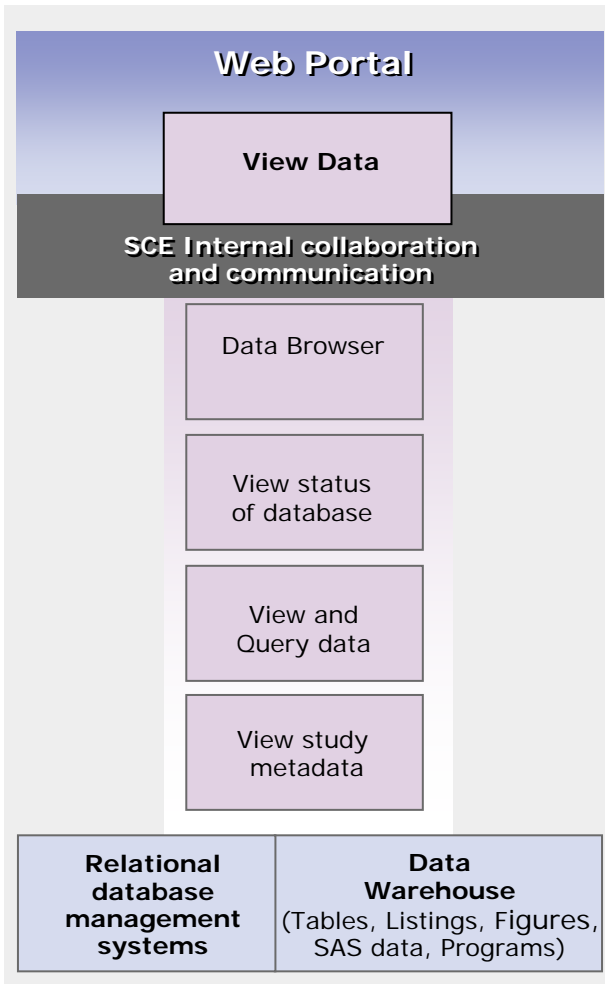
- Less programming
- Direct table specification
  - Less chance for error
  - Reduced validation costs

# 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



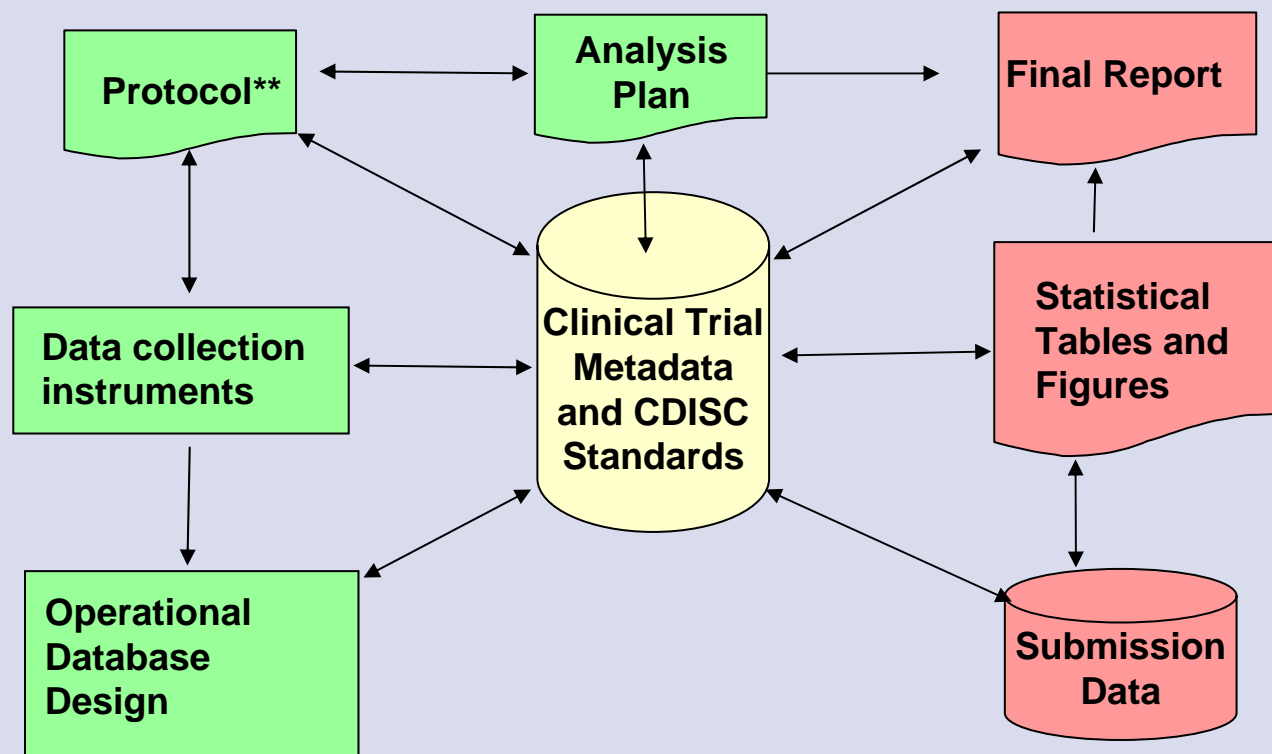
**SDTM data is not readily usable in statistical software nor for viewing in context of other data**

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

# Advantages of a Statistical Computing Environment

- Productivity enhancement
- Facilitates Training
- Manage work objects
  - Analysis files
  - Programs
  - Reports
- Provides process documentation
- Study archival mechanism
- Compliance with Part 11
  - Audit trails
  - File security
  - Version control

# The New World: metadata and data standards will drive clinical trial processes



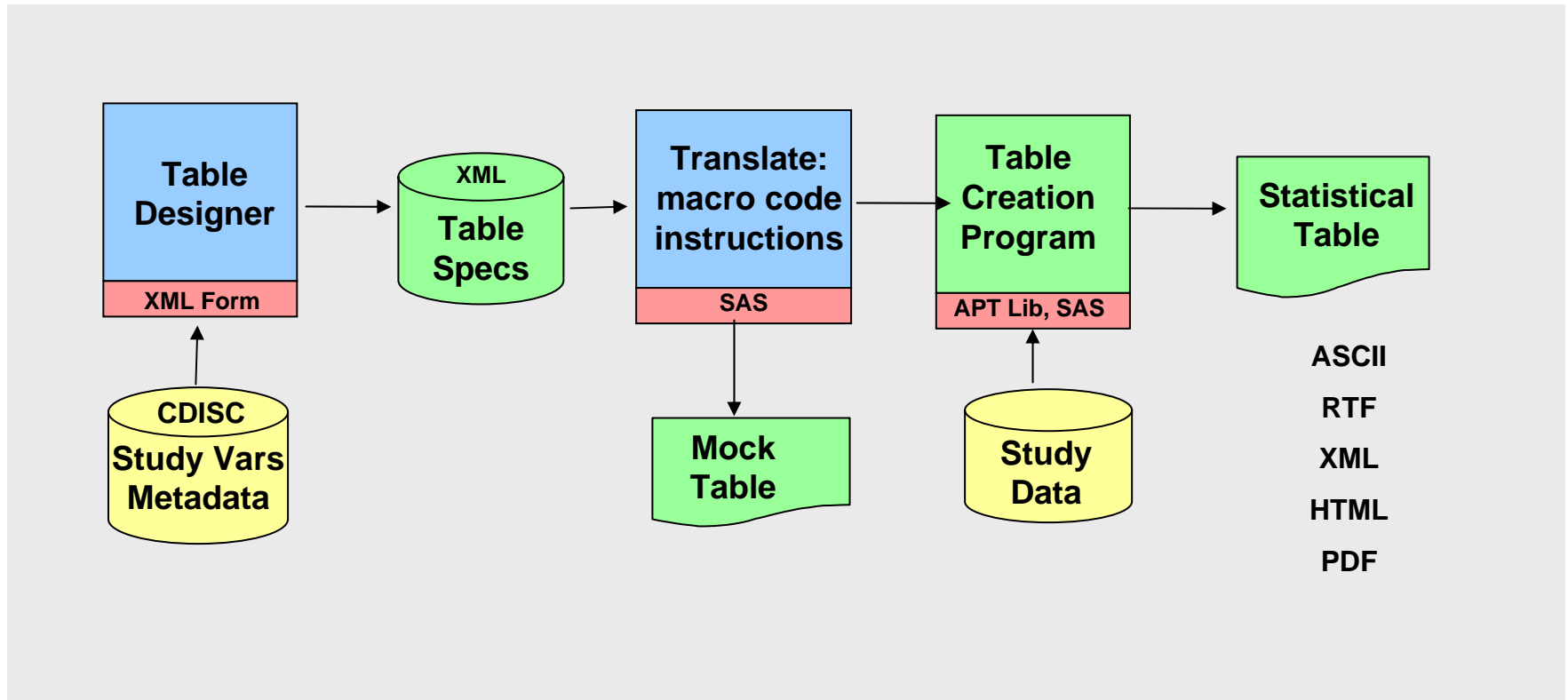
**Publication of a standard, machine-readable model for protocol representation that will facilitate interchange of metadata among systems and stakeholders**

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

Consider the possibilities:

- A basis for developing 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, analysis plan and statistical reports

# Process Flow Automation for Creation of Statistical Tables



Reference: Hopkins & Collins (2005) "Statistical Table Specification and Automatic Code Generation Using XML". PharmaSUG 2005 Proceedings, <http://www.pharmasug.org/2005/papers/ad16.pdf>.

# Electronic Regulatory Submissions

## 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\03.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 expects to receive multiple types of data files, documentation, and programs – the whole statistical environment**

# The FDA Needs *Reproducible Research*

## *Integrate statistical thought and implementation into one document*

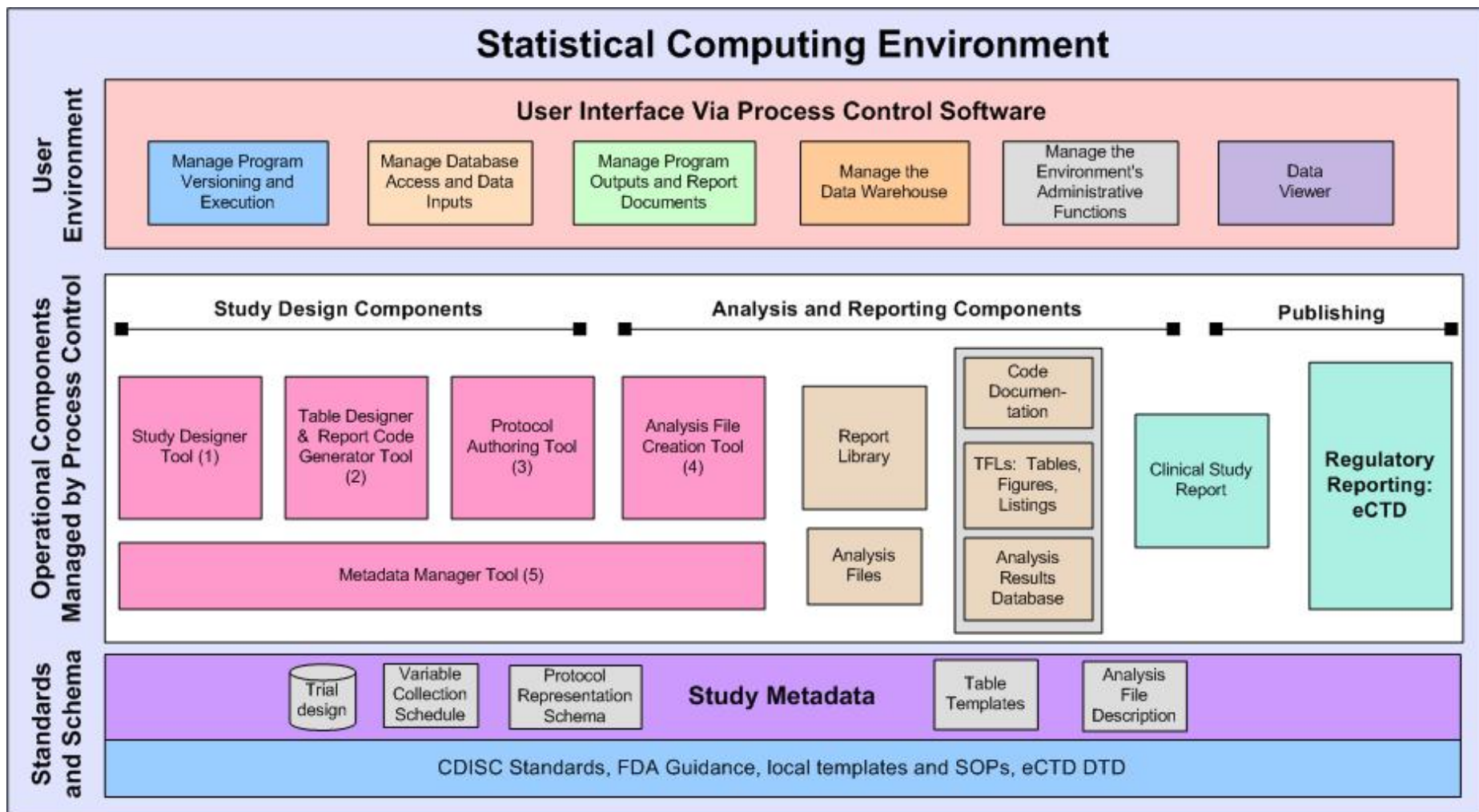
- “*Literate Statistical Practice*” seeks to combine text, statistical programs and code output (tables and figures) into a single document.
- Academics have recognized this need and have made some prototypical efforts. Sweave (combines S-Plus and LaTeX): a system for mixing text and S-Plus code for automatic document creation.
- We need a mainstream system for creating literate statistical documents.

Ref: Rossini & Leisch (2003) “Literate Statistical Practice”. UW Working Paper Series, <http://www.bepress.com/uwbiostat/paper194>.

# Tools to support reproducible statistical analysis

- Statistical computing environment
- Protocol and SAP authoring tool
- Automatic creation of table templates (“mock tables”)
- An interchange medium for a statistical computing environment
- Metadata management
  - Data standards
  - Analysis metadata
- Table specifications
- Enhanced table creation production software
- Smart report authoring environment integrating stat software and documents

# Process Automation Suite Components within a Clinical Trials Statistical Computing Environment



# Summary: Tools for the new processes

**New data standards provide an opportunity to structure traditional statistical processes for better communications within drug development teams and regulatory agencies. New tools are needed to create the future possibilities.**

- 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.
- Process automation is possible if standards are available. Well-defined processes can communicate through XML to create workflows.
  - Protocol authoring
  - Analysis specifications
  - Analysis creation
  - Creation of compound documents designed for analysis reproducibility