genzyme

Standards-based, Metadata-driven Statistical Computing Environments

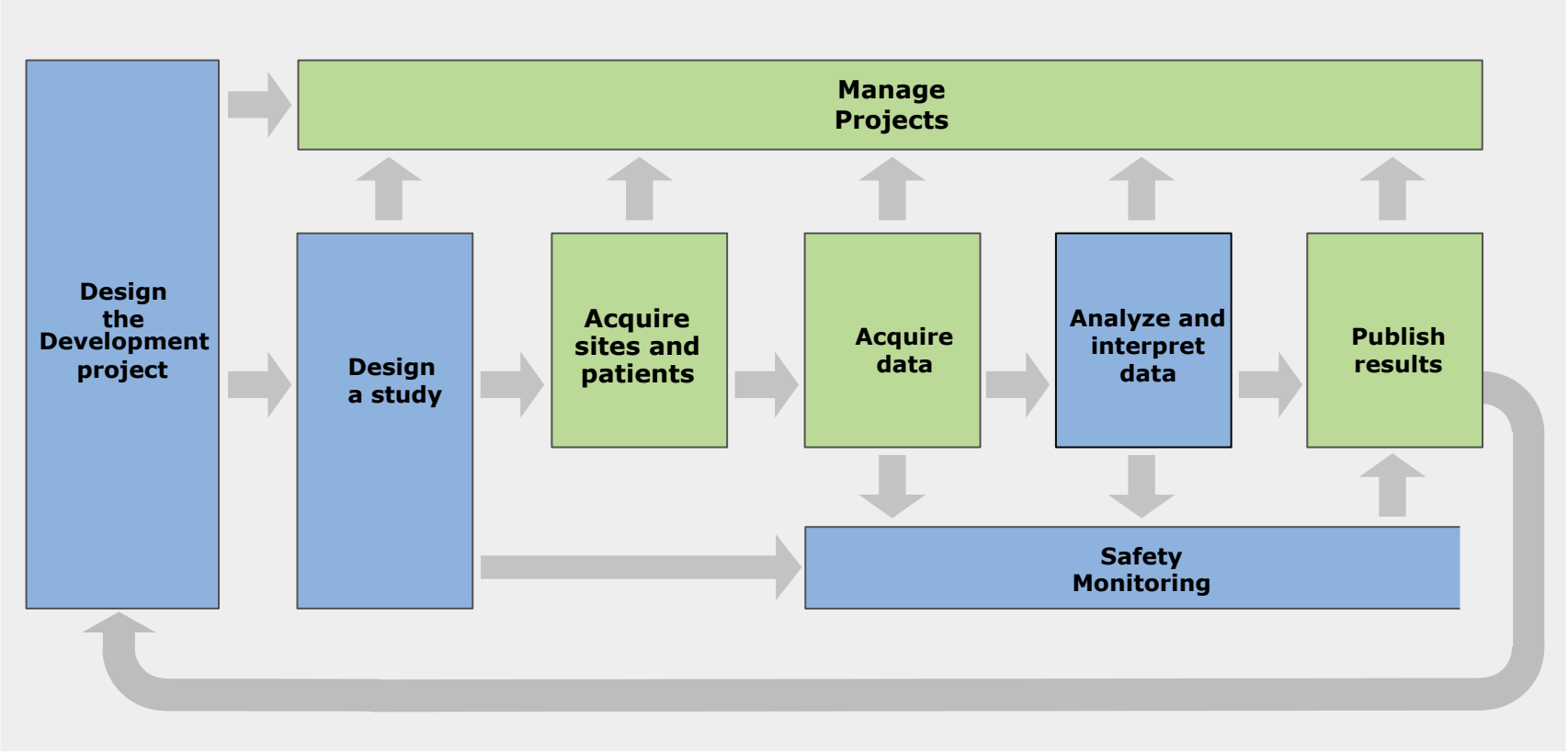
Sue Dubman

September 28, 2010
PhUSE, Cambridge, MA

Outline

- Background and Rationale
- Strategic Vision
- Standards-based, Metadata-driven Architecture
- Acquisition and Implementation
- Lessons Learned

The Clinical Development Process



Source: Hopkins, Duke, Dubman, *"Statistical Computing Environments and the Practice of Statistics in the Biopharmaceutical Industry",* Drug Information Journal, Vol. 44, 2010.

genzyme

Statistician's Role in Clinical Development

Clinical Research Assets

Functional Roles	Protocol	CRF	Analysis Plan	Study Report	ISS, ISE, etc	Label	Re-Use
Clinical Dev	X	X	X	X	X	X	?
Monitoring	X	X		?			
Data Mgmt	X	X	?				
Statistics	X	X	X	X	X	X	X
Medical Writing	?			X	X	?	?
Regulatory	X	?	?	X	X	X	

X = always involved in creating this asset
? = amount of involvement with this asset may vary by organization

genzyme

Duke, 43rd DIA Annual Meeting, 2007.

Good Statistical Practice leads to credibility of results: reproducible, transparent and validated analyses

Good Statistical Practice

Science

- Protocol
  - **Minimize bias**
  - **Maximize precision**
- Analysis plan
- Presentation of results

+

Operational Processes

- Controlled statistical environment
- SOPs
- Productivity tools
- Data standards
- Training

→

Credibility

Results

- Reproducible research
- Transparent and efficient processes
- Validated analysis
- Data integrity assurance

Source: Hopkins, 2006.

Drivers for creation of controlled statistical environments

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

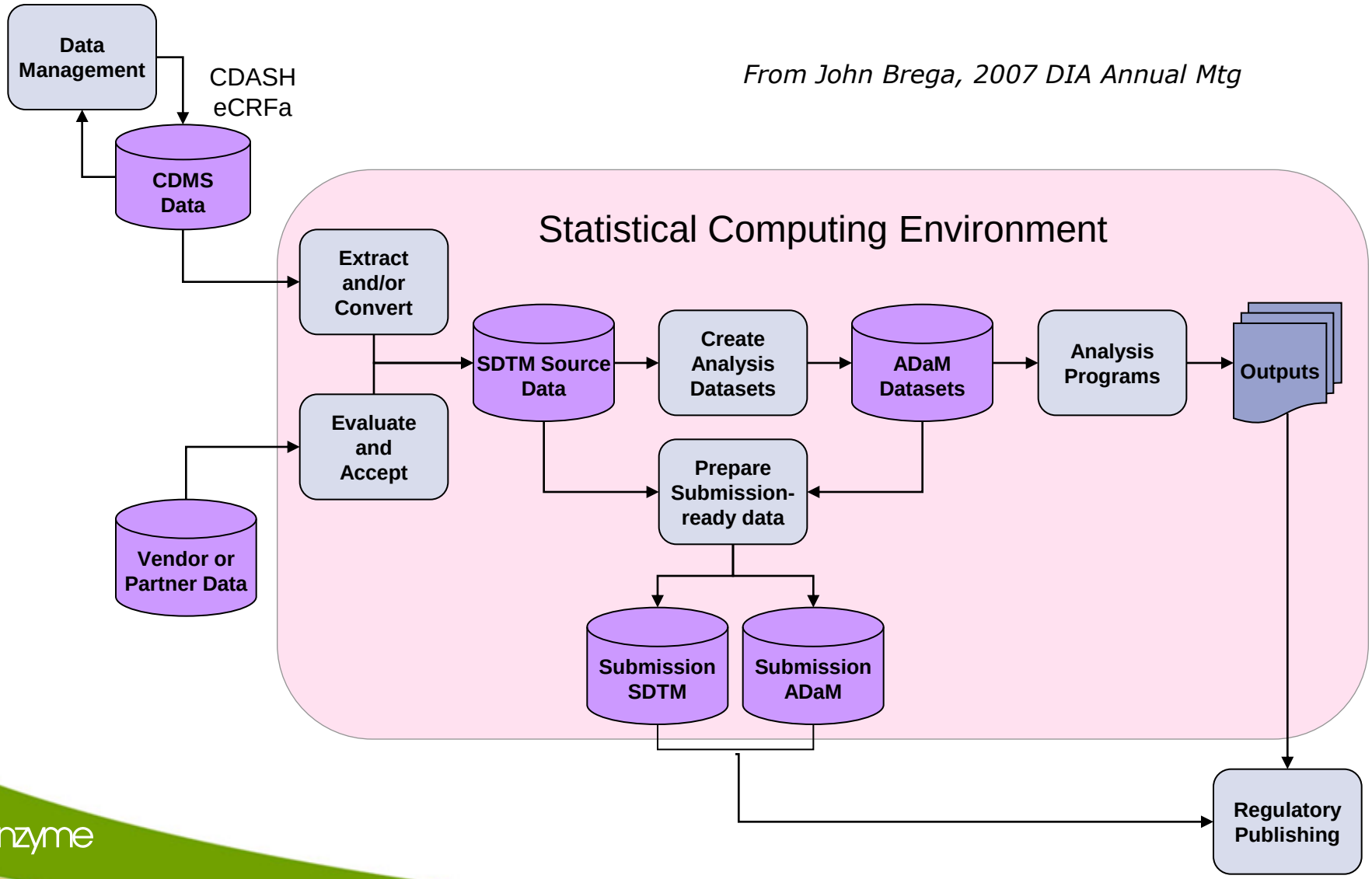
- **FDA Regulations/Guidelines**
  - ICH E9: Statistical Principles for Clinical Trials
  - 21 CFR Part 11 & Scope and Application Guidance
  - "Computerized Systems Used in Clinical Trials" (April 1999; Sept 2004)
  - eCTD specification
- **Standards**
  - CDISC: ODM, SDTM, ADaM, CDISC/HL7 Protocol Representation
  - Statistical analysis plans (ICH E3, CDISC/HL7 subcommittee)
- **Increasing Complexity**
  - **Technologies**: e.g. XML, Internet
  - **Outsourcing and data sharing** (CROs, Labs, DMCs, Partners, FDA, …)
- **Cost of validation** – what can be done to minimize cost?
- **New tools** – natural response to the FDA Critical Path Initiative of 2004

Statistical Computing Environment (SCE) Vision

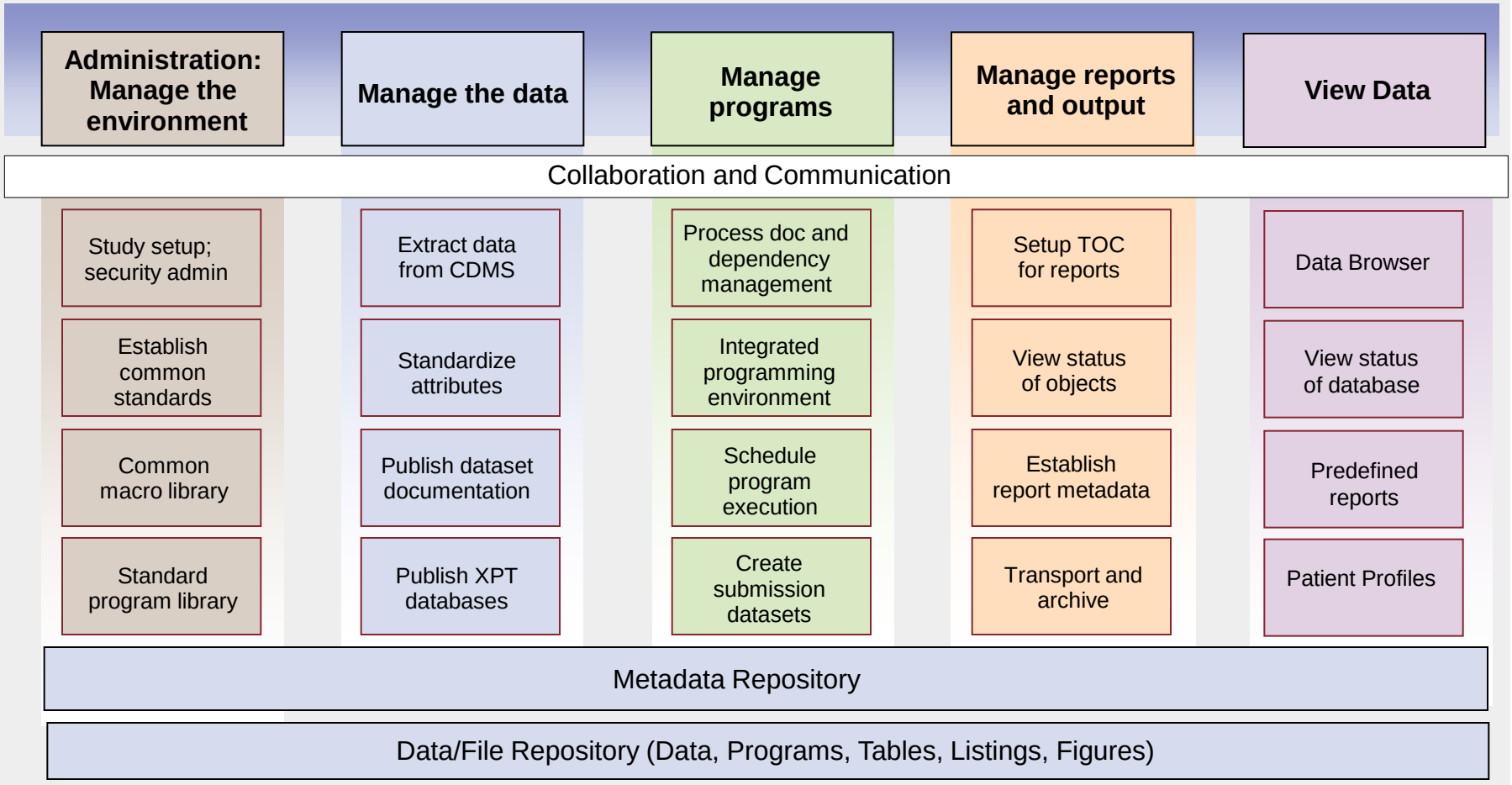
A system that provides a foundation for documenting rigor in the analysis and reporting of clinical trial results while *increasing productivity and quality*.

genzyme

Domain of Statistical Computing

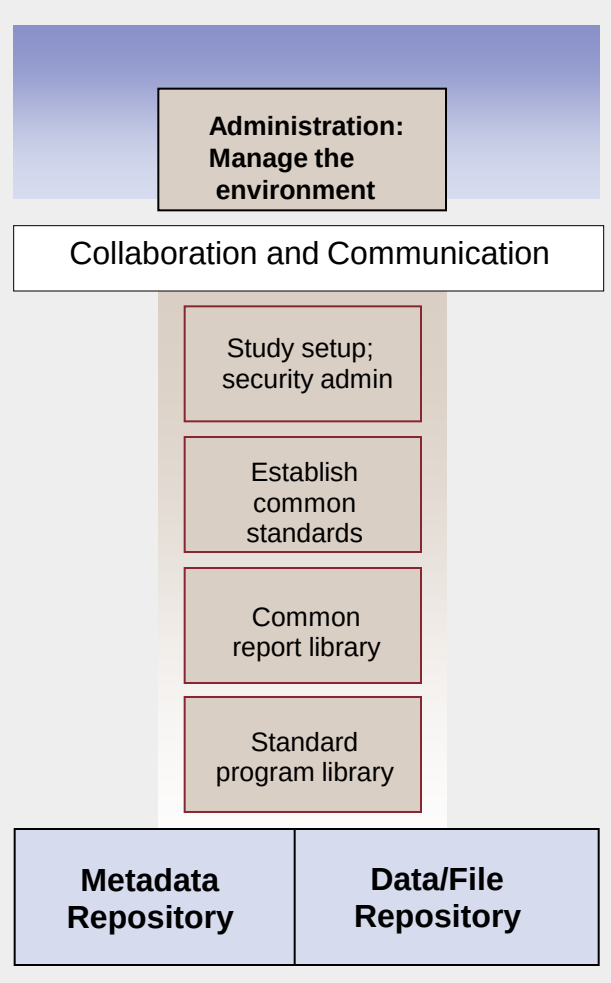


SCE: Conceptual Components



Adapted from Hopkins, 2006

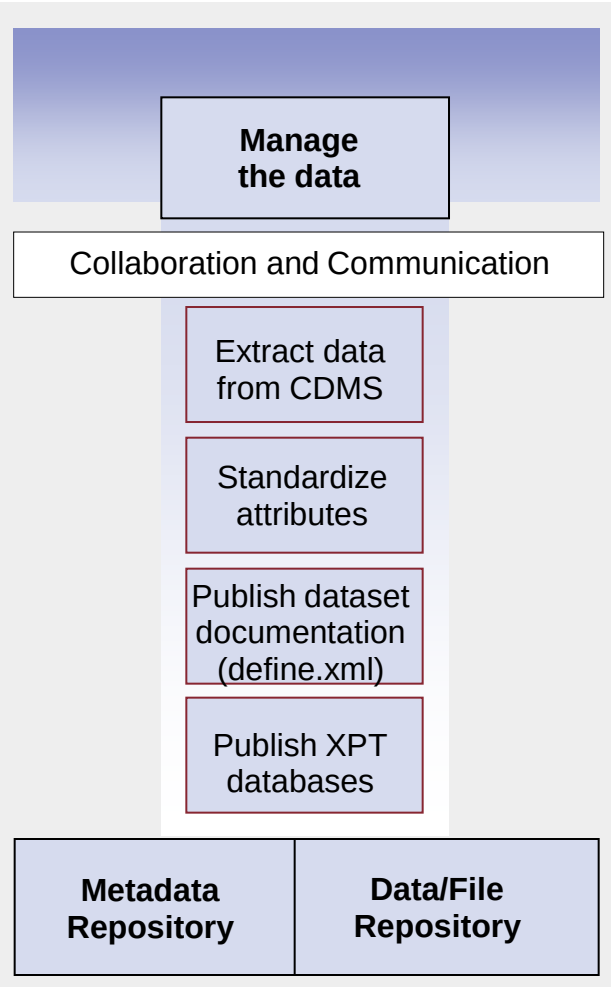
Manage the Statistical Environment



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

genzyme

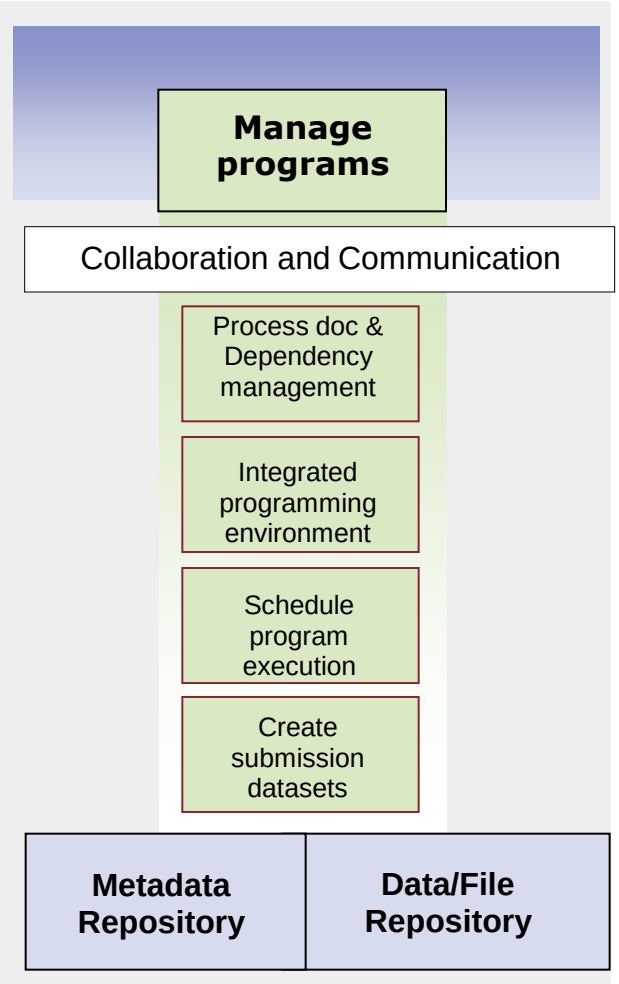
Manage the Data Repository



The data/file repository is a single place for storing documents, data, programs, and output

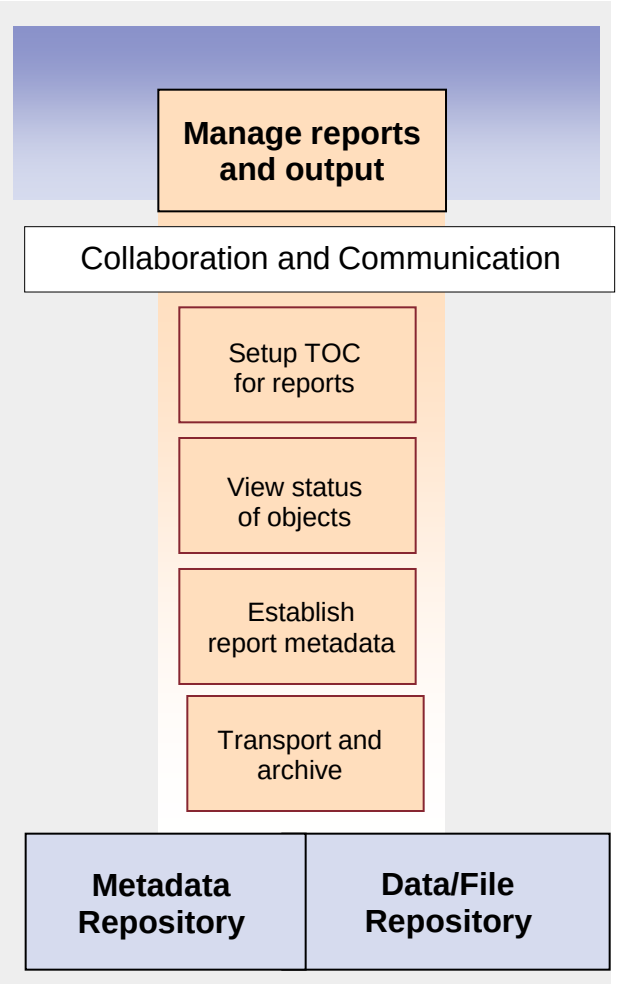
- Support for import of SDTM data from clinical data management or vendor
- Provide data security and audit trails for the data warehouse
- Store data, logs, programs and output. Handle these components as related objects.
- Store statistical analysis files

Manage Analysis Programs



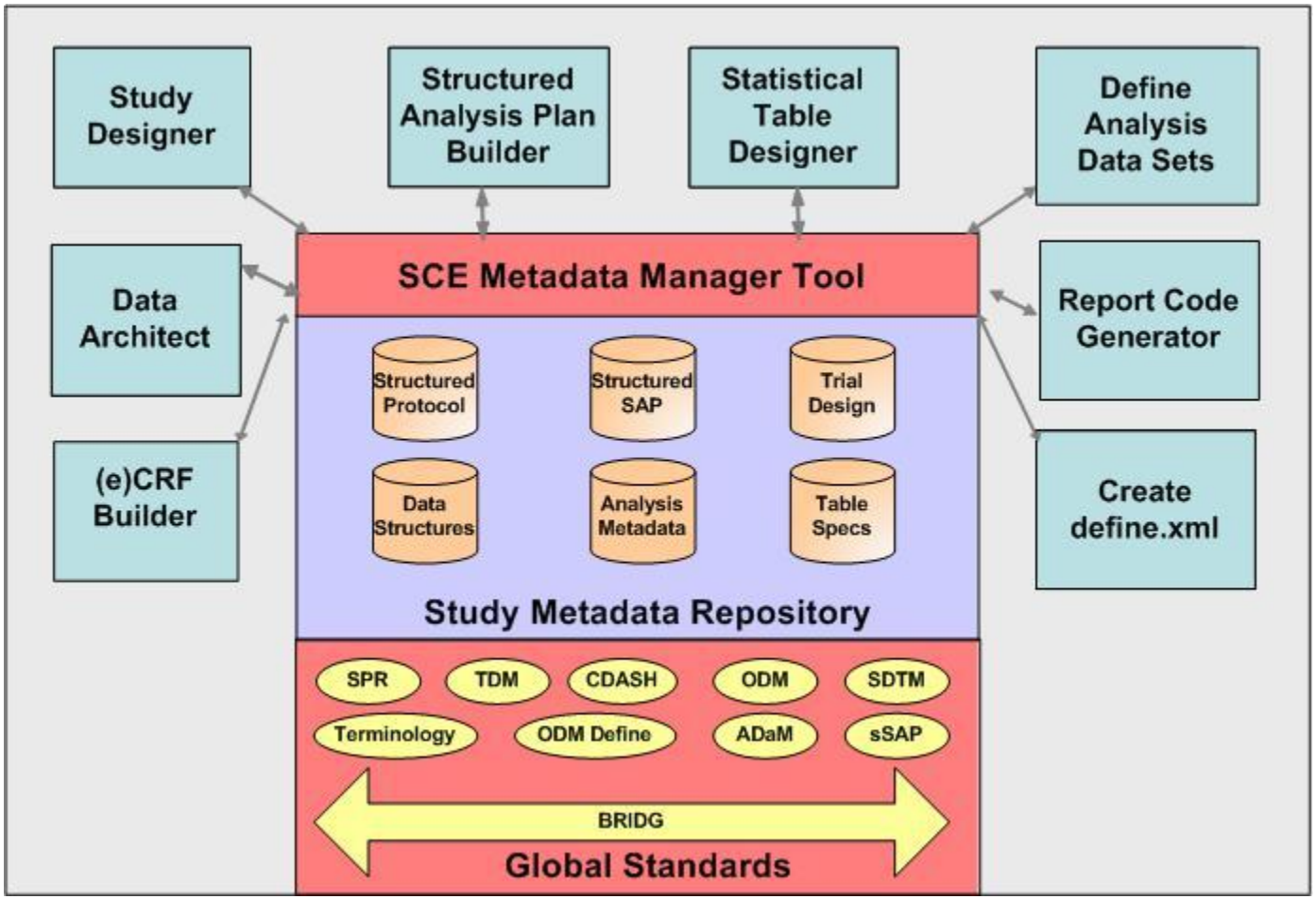
- Code management: Check-in / check-out
- Audit trails
- Batch processing
- Dependency management
- Extensible: supports multiple tools
- Flexible: allows program execution outside the environment for early phases of program development

Manage Reports and Output



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

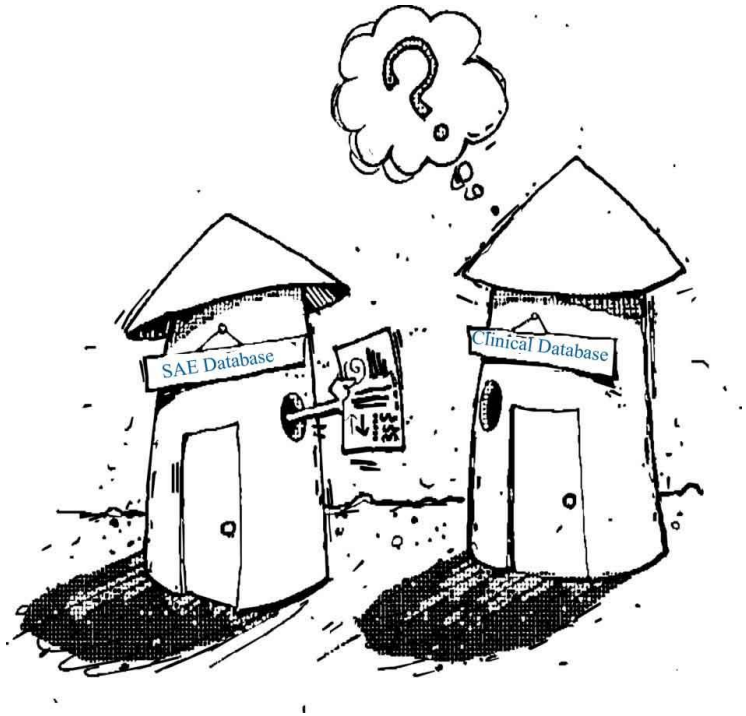
Metadata at Center of Statistical Workflow



genzyme

Why is Metadata Needed?

- Metadata answers the who, what, when, where, why, and how about every facet of the process and study data
- Needed to ensure consistency in use and meaning of content
- The SCE itself is a major producer of metadata needed to manage statistical processes and work flow



genzyme

Metadata-driven SCE

- Every object that is managed has metadata, so it can be determined how the pieces are hooked together.
- Metadata consist of both global and study-specific data element definition standards, other content information, and process information.

genzyme

A Metadata Repository is a Mandatory Component of an SCE

A metadata repository isn't a fancy spreadsheet to store standards!

It's a requirement for standards-based automation which is key to ROI from standards implementation

- Manage multiple versions of single data element
- Capture the use of each standard
- Select different "sets" of standard objects to use for a particular program
- Manage workflow and change control
- Facilitate cross-functional and cross-organizational use of standards
- Facilitate communication between systems
- Provide impact analysis, inheritance, etc. to manage changes in standards

genzyme

Acknowledgements: Dana Soloff, Genzyme

Data Attributes ~5000 per Study*
Each attribute is manually typed – THREE TIMES!

: Work required to add one new analysis population variable

Documentation

Variable Name: RESCPOP

Valid Values: Y, N

Label: Rescue Population

Format: $yesno.

Type: Char

SAS Program

ATTRIB

RESCPOP

Length = $1

Format = $yesno.

Label = “Rescue Population”

QC Program

ATTRIB

RESCPOP

Length = $1

Format = $yesno.

Label = “Rescue Population”

Study Specific Derivation Description

Study Specific Derivation SAS Code

Study Specific Derivation SAS Code

*25 SDTMs, 20 ADaMs, 20 variables each, 5 attributes per variable
+ 500 elements of controlled terminology

genzyme

Acknowledgements: Dana Soloff, Genzyme

Manual QC Required to Ensure Consistency Within a Study…

Study 111

Documentation

Variable Name: RESPOP

Valid Values: Y, N

Label: Rescue Population Set

Format: $yesno.

Type: Char

Study Specific Derivation Description

SAS Program

ATTRIB

RESCPOP

Length = $1

Format = $yesno.

Label = "Rescue Population Set"

Study Specific SAS Derivation

Changes to datasets require regenerating TLFs and QC

genzyme

Acknowledgements: Dana Soloff, Genzyme

…And Across Studies

Study 111 created a numeric variable - study 222 created a character

Study 111

Study 222

Documentation

Variable Name: RESCPOP

Valid Values: Y, N

Label: Rescue Population Set

Format: $yesno.

Type: Char

Study specific derivation description

SAS Program

ATTRIB

RESCPOP

Length = $1

Format = $yesno.

Label = "Rescue Population Set"

Study Specific Derivation: SAS Code

Documentation

Variable Name: RESCPOP

Valid Values: 1,0

Label: Rescue Population Set

Format: yesno.

Type: NUM

Study Specific Derivation Description

SAS Program

ATTRIB

RESCPOP

Length = 8

Format = yesno.

Label = "Rescue Population Set"

Study Specific Derivation SAS Code

genzyme

Acknowledgements: Dana Soloff, Genzyme

Acquisition & Implementation

- Use cases → Requirements → RFP
- Vendor Selection → Pilot Study
- Modify/develop SOPs to accommodate SCE processes
- Rollout
  - –Just-in-Time Training
  - –Dedicated IT and Stakeholders for conversion support and rollout
- Production

Acknowledgements: Linda Collins, DIA 2009

Process Realignment

- An SCE will not be able to reach its full business value if implemented without some level of established and enforced statistical programming business process.
- Expect change:
  - Expect to revise procedures
  - Expect improved control of processes and documentation
  - Expect improved tracking of project
  - Expect better quality work!
- Implementing SCE in new ways will yield new results

Implementing SCE in old ways will yield old results

Getting used to the new world

- Best practices for check in/check out
  - Version only meaningful changes
- Best practices for assigning status flags
  - What does ‘draft’ mean?
- Managing the culture
  - Avoid ‘gotcha’ attitude
- Surprises:
  - We can tell what changed when and why!
  - Executable metadata is a beautiful thing

Contacts

- Sue Dubman

  Senior Director, Genzyme, Inc.

  sue.dubman@genzyme.com

genzyme