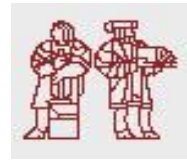




The MIT Information Quality Industry Symposium, 2007



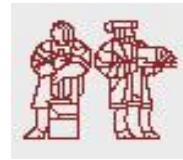
Challenges and Opportunities in Clinical Trial Data Processing

Vadim Tantsyura, Olive Yuan, Ph.D. Sergiy Sirichenko
(Regeneron Pharmaceuticals, Inc., Tarrytown, NY)



Introduction

- “The review and approval of new pharmaceuticals by federal regulatory agencies is contingent upon a trust that the clinical trials data presented are of sufficient integrity to ensure confidence in the results and conclusions presented by the sponsor company.” (Society for CDM, Charter of the Committee for Standards for GCDMP, 1998.)
- Risks Associated with Data Quality
 - Audits / Due diligence / Potential partners
 - Analysis of efficacy
 - Integrated Analysis of Safety
 - FDA inspections / Refuse To Files



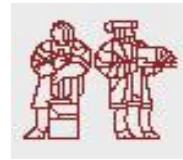
Definition of DQ in Clinical Trials

SCDM adapted the IOM definition:

- (sufficient) “quality data is data that support conclusions and interpretations equivalent to those derived from error-free data” (Institute of Medicine, Roundtable Report, 1999)
 - What DQ means is fairly clear
 - How to get there is debatable (many ways)
- Steps to make this definition operational
 - Understand sources of errors
 - Identify errors through review/inspection/audit
 - Use inspection results to measure data quality
 - Assess impact of data quality on conclusions drawn from the trial (statistician’s responsibility)



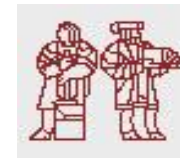
The MIT Information Quality Industry Symposium, 2007



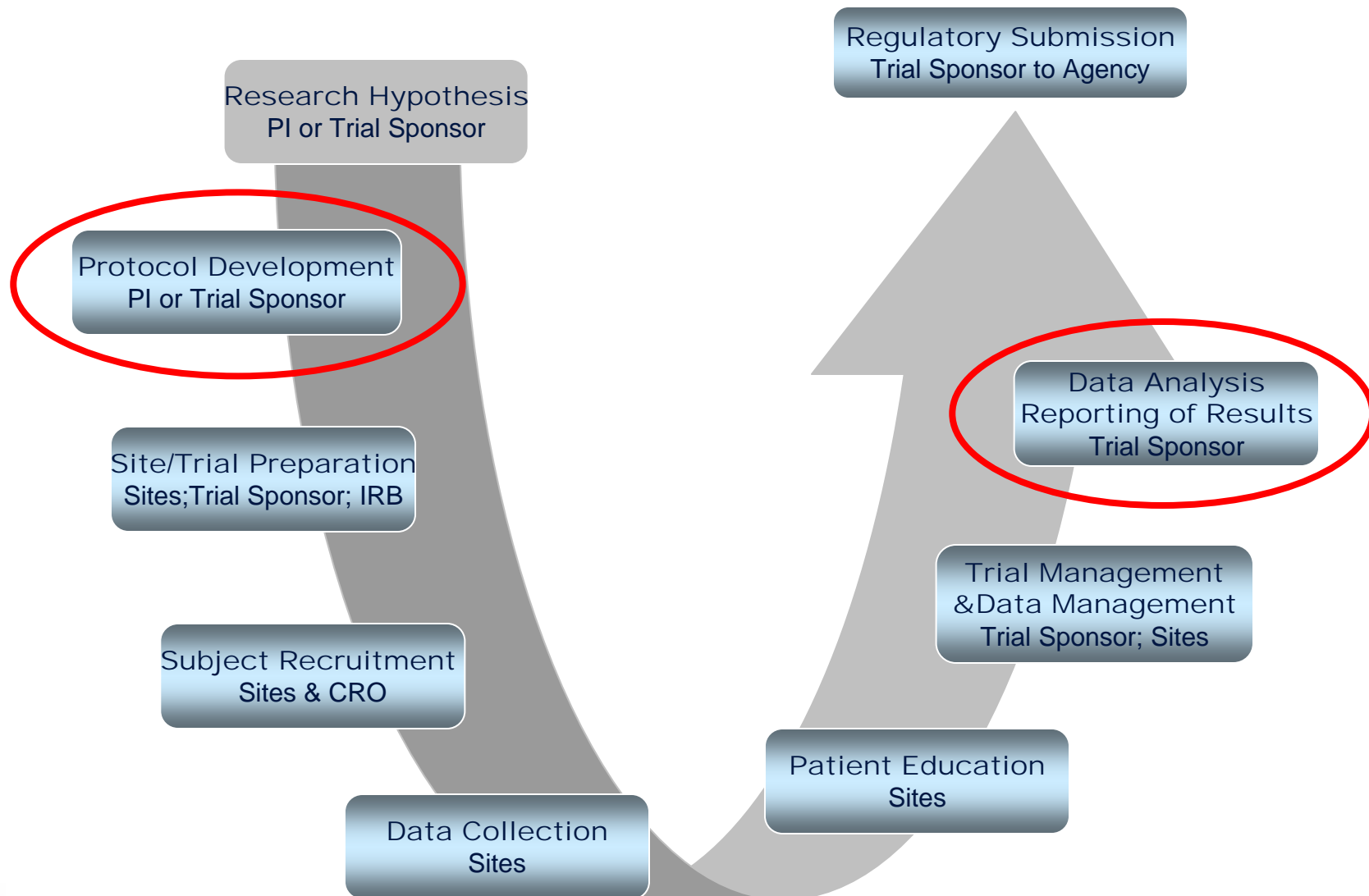
Agenda:

1. Overview of Clinical Trial (CT) Data Collection / Cleaning Process
 - Definition DQ in CT
 - CT is a multi-step process: quick overview
 - The sources of data errors
 - Steps in correcting errors and controlling variability
 - Assuring DQ:
 - **Industry**
 - **FDA**

2. Data standardization
 - CDISC SDTM
 - CDISC CDASH

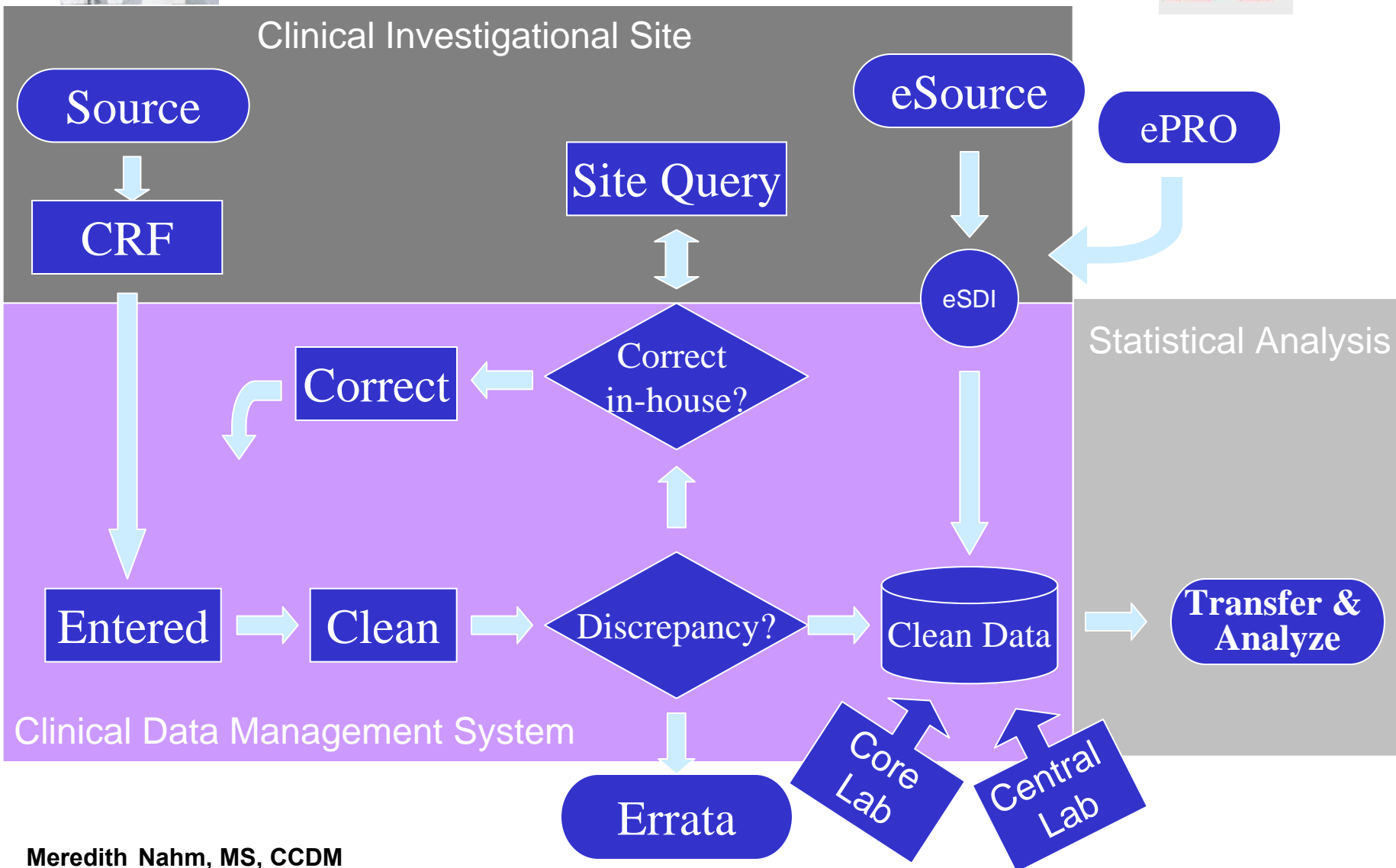


Today's Clinical Trial





The MIT Information Quality Industry Symposium, 2007

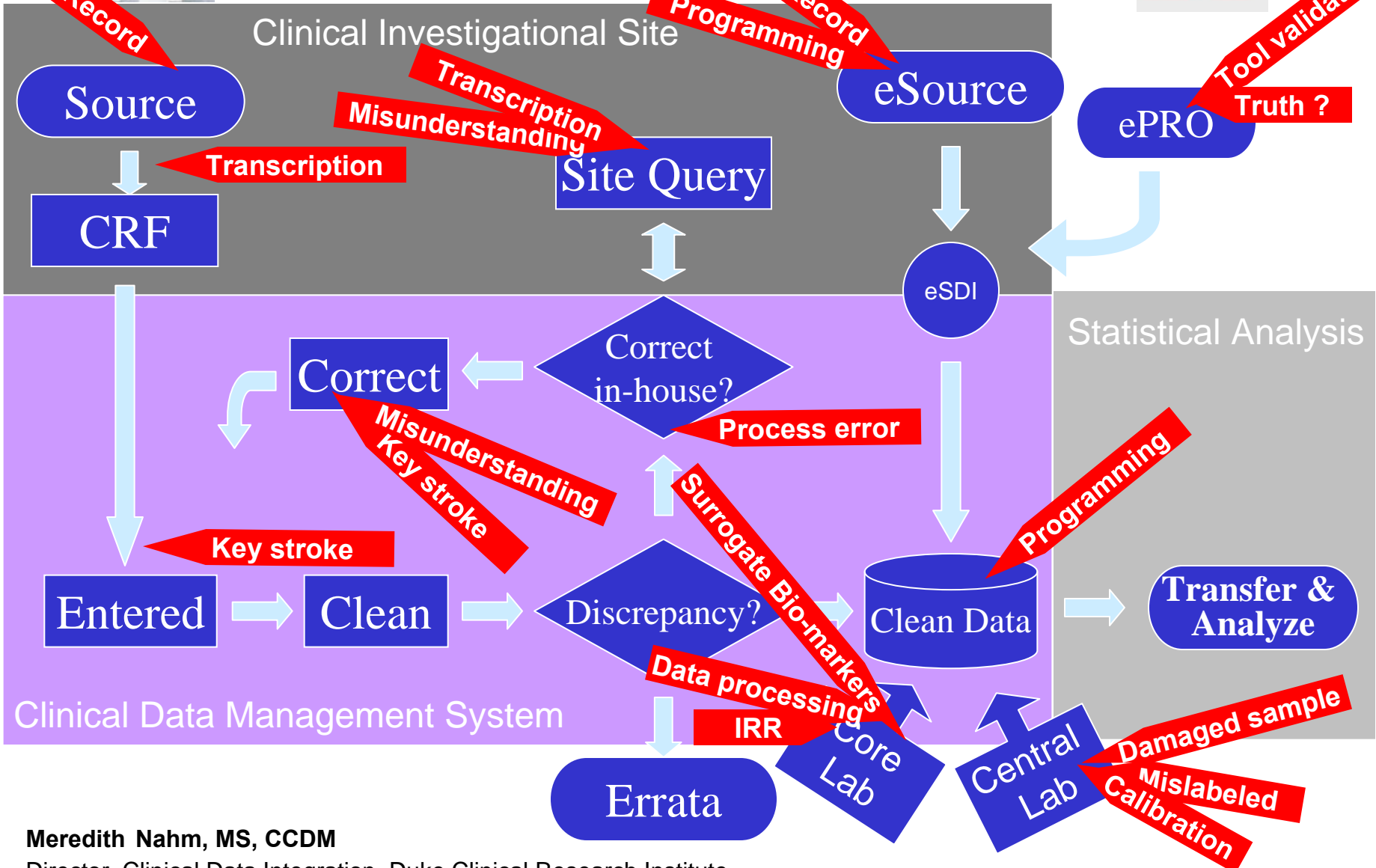


Meredith Nahm, MS, CCDM

Director, Clinical Data Integration, Duke Clinical Research Institute

Author, GCDMP Measuring Data Quality and Assuring Data Quality Sections

The MIT Information Quality Industry Symposium, 2007



Meredith Nahm, MS, CCDM

Director, Clinical Data Integration, Duke Clinical Research Institute

Author, GCDMP Measuring Data Quality and Assuring Data Quality Sections



The Sources of Data Errors

Patients/Subjects

- Forgotten or misremembered details
- Omission of data considered to be of no consequence
- Untruthfulness/reluctance to provide a “wrong answer”
 - **“Yes doctor, I’m taking the pills exactly as you directed”**
- Misunderstanding of a question leads to inappropriate response
- Responding to questions when nervousness
 - **~5% of the time patients will report their own age incorrectly**



The Sources of Data Errors (cont'd)

Site Personnel

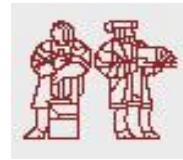
- Omissions: forgotten details & data considered “of no consequence”
- Misunderstanding of question
- Misinterpreted handwritten notes
- Errors in extraction from source document to CRF or eCRF (commonly called “transcription errors”)
- Incorrectly performed procedures
- Collecting data outside a required time window
- Human miscalculation of a count, total, etc...
- Measurement error
- Guessing/assuming when factual info is not readily available
- Lack of inter-rater reliability (cases where there is subjectivity and more than one individual is assessing the data)



The Sources of Data Errors (cont'd)

Study Sponsor & CRO Personnel:

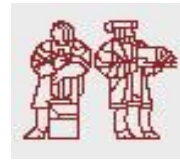
- **Errors in data processing**
 - Misinterpreted handwritten values
 - Errors in keying
 - Miscalculation of a derived item
 - Providing misleading or confusing instructions or queries to site personnel – leading them to introduce new/additional errors
 - Inappropriate use of “Self-Evident Corrections”
 - Using “leading language” to suggest “better” values
 - Errors due to data integration issues
- **Errors in database** (e.g. data stored in wrong place)



The Sources of Data Errors (cont'd)

Deliberately introduced errors

- “Little white lies”
 - **Common example is entry criteria rounding**
 - **SBP \geq 165 and DBP \geq 90 is required for entry into the study**
 - **BP is measured as 163/91 but recorded as 165/91**
 - **Studies show that significantly more patients “barely meet” than “barely miss” entry criteria**
- Fraud
 - **Subject doesn't actually exist**
 - **Culture from 1 subject is reused to qualify 2nd subject into the study**

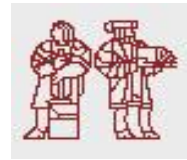


Source Data Verification (SDV) Findings

- CRF data is verified/revised based on source notes
- Common to find limited source documentation for:
 - Physical exam, medical history, concomitant medication
 - Anything the site directly records onto CRFs during subject visits
- Less commonly (but not rare), source docs may be scant for:
 - Primary endpoints
 - Dosing of study drugs
 - Safety assessments including AEs
- Notations in the source documents that do not make sense
 - Transcriptions errors & Inconsistencies
- Additional information not recorded on the CRF
 - Repeat labs
 - Additional measurements not required by the protocol
 - Notes about non-study related visits if site is primary care provider

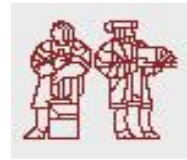


The MIT Information Quality Industry Symposium, 2007



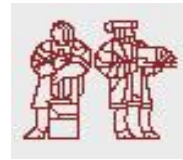
SDV Cannot Find:

- Data that exists in patient charts located with other care providers
 - **Major limitation for subjects recruited via referral or advertising**
 - **Interestingly, these subjects are often perceived as “very clean”**
- Information the patient didn't report or misreported
 - **Cultural factors have a big impact on what subjects will report**
- Data the subject reported, but site staff considered of no consequence
 - **Some still fail to record all adverse “events” in favor of selectively reporting those they consider adverse “reactions”**
- Some forms of fraud
- Errors in the data that the monitor doesn't review



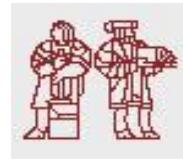
Other Data Cleaning Steps

- Impossible values (inconsistent with life) are usually updated
- Medical inconsistencies are usually resolved
- Missing values are confirmed missing or provided
- Extreme values are confirmed or replaced with more acceptable values
- Many changes to items included to facilitate CDM or statistical processing are modified
- Inconvenient/unexpected data is removed, modified, or explained away



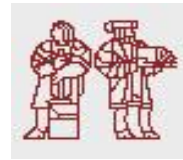
Unintended Consequences

- Bias: systematically focusing on extreme values when as many errors are likely to exist in the expected range
- Bias: selectively prompting to modify or add non-numeric data to make it “right” – EVEN when the cleaning is fully blinded and well intended
- Before selective cleaning, the data may be flawed,
BUT the flaws are not systematically concentrated
- Reduced variance ⇒ increases Type I Error Rate
 - Risk of making incorrect inferences
 - Finding statistically significant differences due to chance alone
 - Failing to find differences in a non-inferiority comparison
- Processing errors are often introduced into the data during cleaning



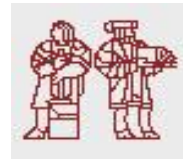
Assuring Clinical Trial Data Validity: Industry Part

- **The complexity of the design and the amount of data collected have important influences on data quality**
 - Design of protocol
 - CRFs
 - Data collection systems
- **Training is critical to ensuring that the protocol is followed correctly and the CRFs are properly completed**
 - Clinical investigator
 - Study personnel
- **Clinical site monitoring (can consume 15 to 30 percent of overall trial costs)**
- **Industry data QA procedures**
 - Assembly of all the data from trial
 - Entry of the information into databases
 - Evaluation of the data for quality
 - Audits of clinical sites



Assuring Clinical Trial Data Validity: The FDA Part

- **FDA data analysis (includes clinical and statistical review)**
 - Checking and verification of data from important analyses submitted by the sponsor
 - Performance of exploratory analyses to answer questions that emerge from the review
- **FDA data QA evaluation**
 - Auditing of CRFs to verify the accuracy of tabulated data
 - Evaluation of follow-ups on reported AEs
 - Verification of primary outcome measure at the CRF level
 - An overall assessment of data quality is developed. If serious questions regarding overall data integrity are not resolved, FDA will not approve the application
- **FDA clinical study audit program**
 - A thorough on-site review of these sites is conducted by trained FDA inspectors. Record keeping, adherence to the protocol, informed-consent procedures, and other aspects of the study are assessed. If objectionable conditions are found, a report (FDA Form 483) is provided to the PI at the conclusion of the audit.
- **FDA enforcement activities**
 - If an investigator found to have serious or repeated problems in performing clinical studies, FDA will take steps to debar the individual from performing trials for regulatory purposes. In cases of fraud, criminal prosecution may be pursued.

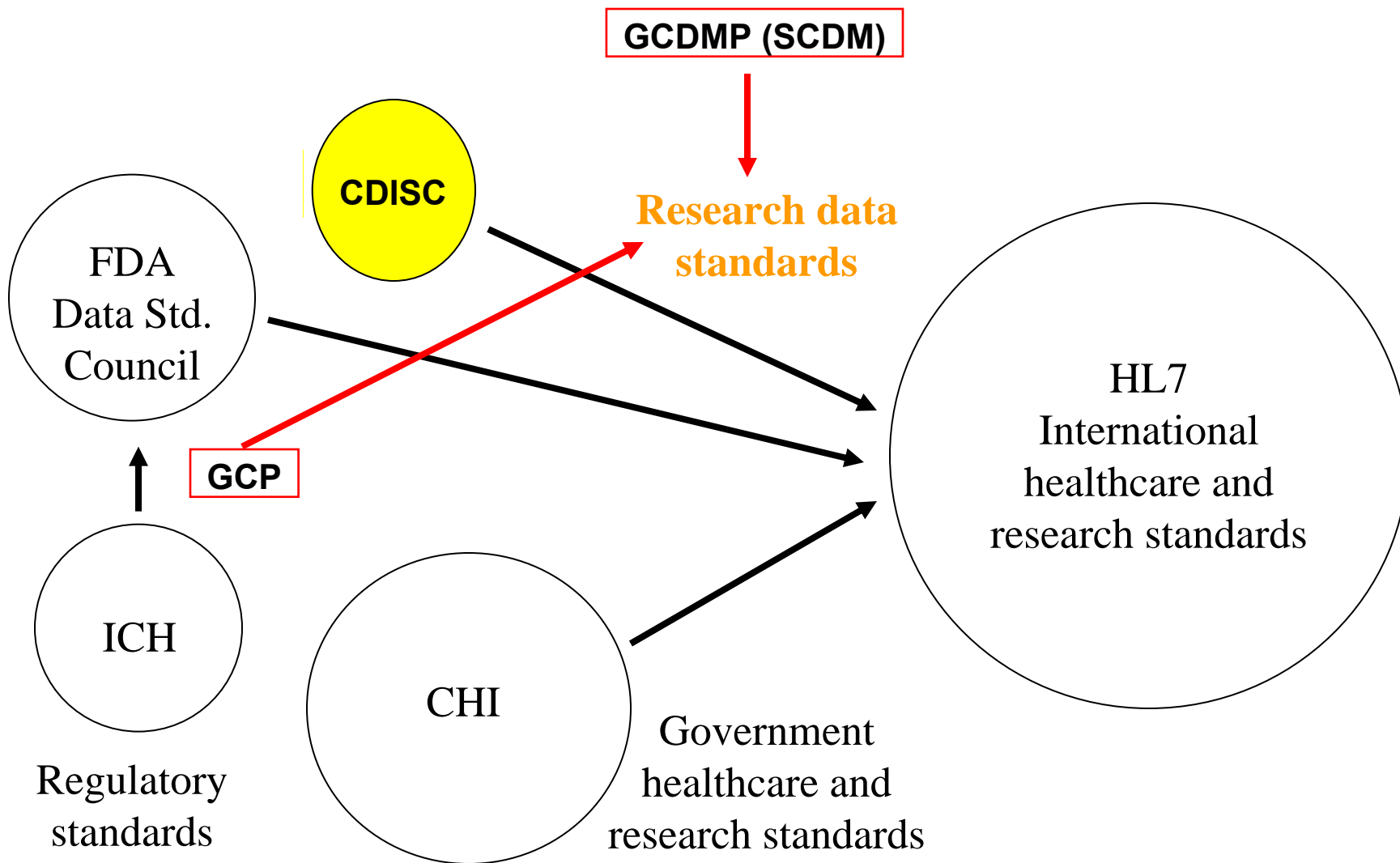


Data Standards – major opportunity and challenge

- Why standardize?
 - Example: [30 AE pages.pdf](#)
- Data Standards (including CRFs) – many types (CDASH, HL7,...)
 - “FDA is serious about CDISC.” (*Dr. R. O’Neill, FDA*)
 - “The importance of a standard for the exchange of clinical trial data cannot be overstated. FDA reviewers spend far too much valuable time simply reorganizing large amounts of data submitted in varying formats. Having the data presented in a standard structure will improve FDA’s ability to evaluate the data and help speed new discoveries to the public.”
-Lester Crawford, Acting Commissioner, FDA (07/21/2004)
- **CDISC** = Clinical Data Interchange Standardization Consortium
 - CDISC **SDTM**
 - CDISC **CDASH**

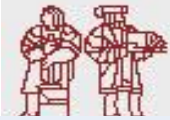


Standards Organization

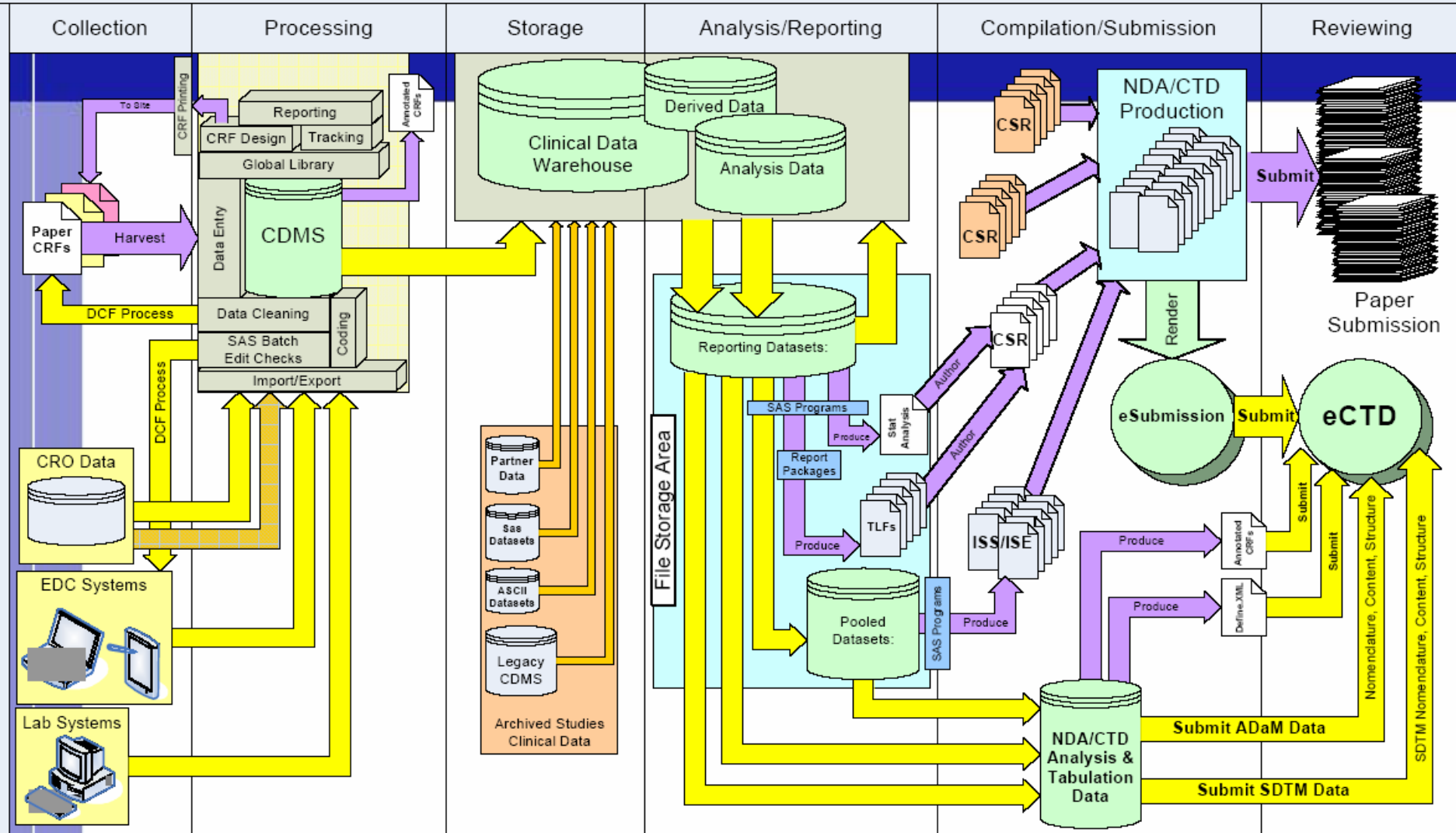




The MIT Information Quality Industry Symposium, 2007



Clinical Information Process Flow – Proposed Future with SDTM

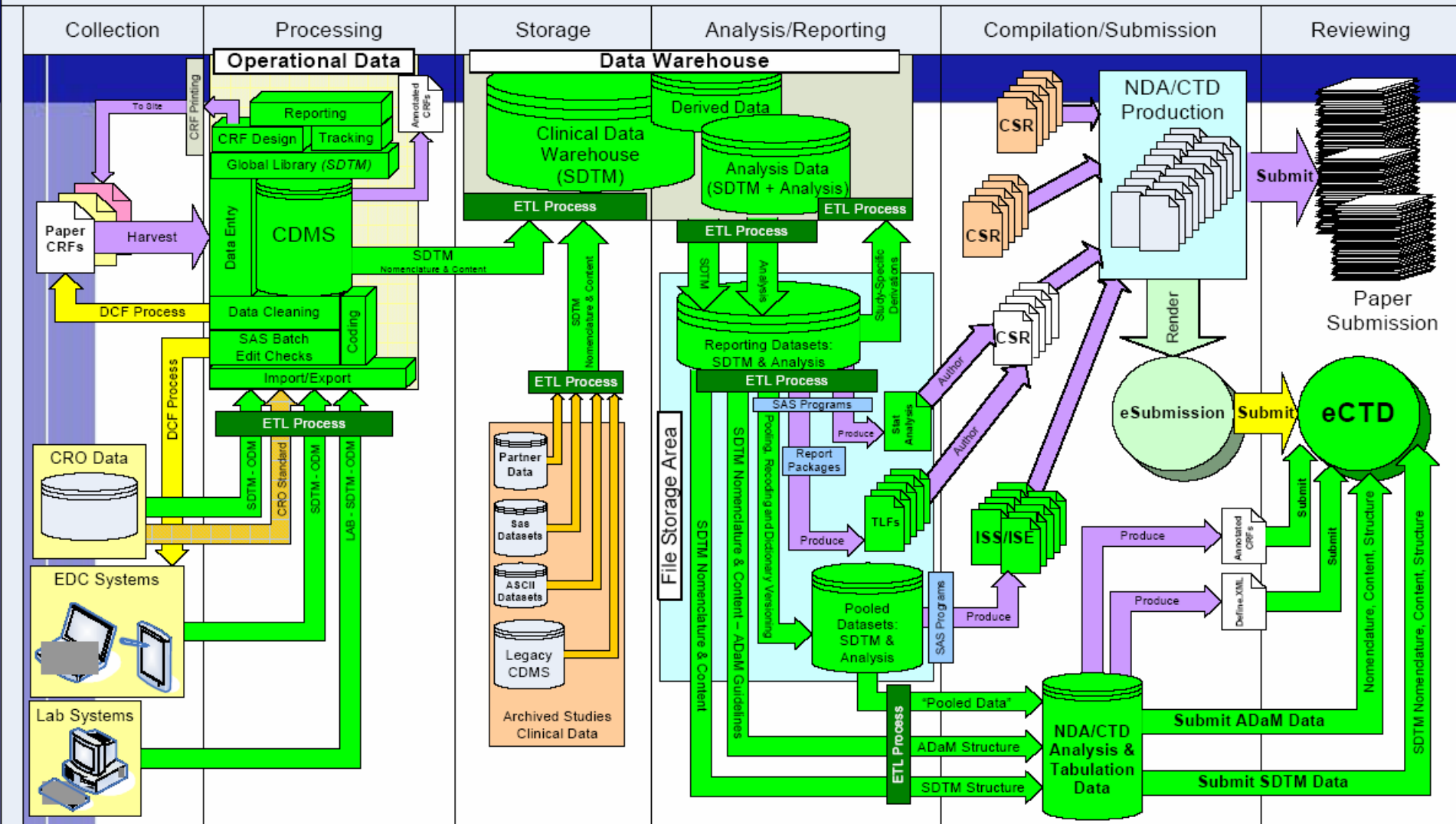




The MIT Information Quality Industry Symposium, 2007



Clinical Information Process Flow – Proposed Future with SDTM



September 26, 2006

Octagon Research Solutions, Inc.

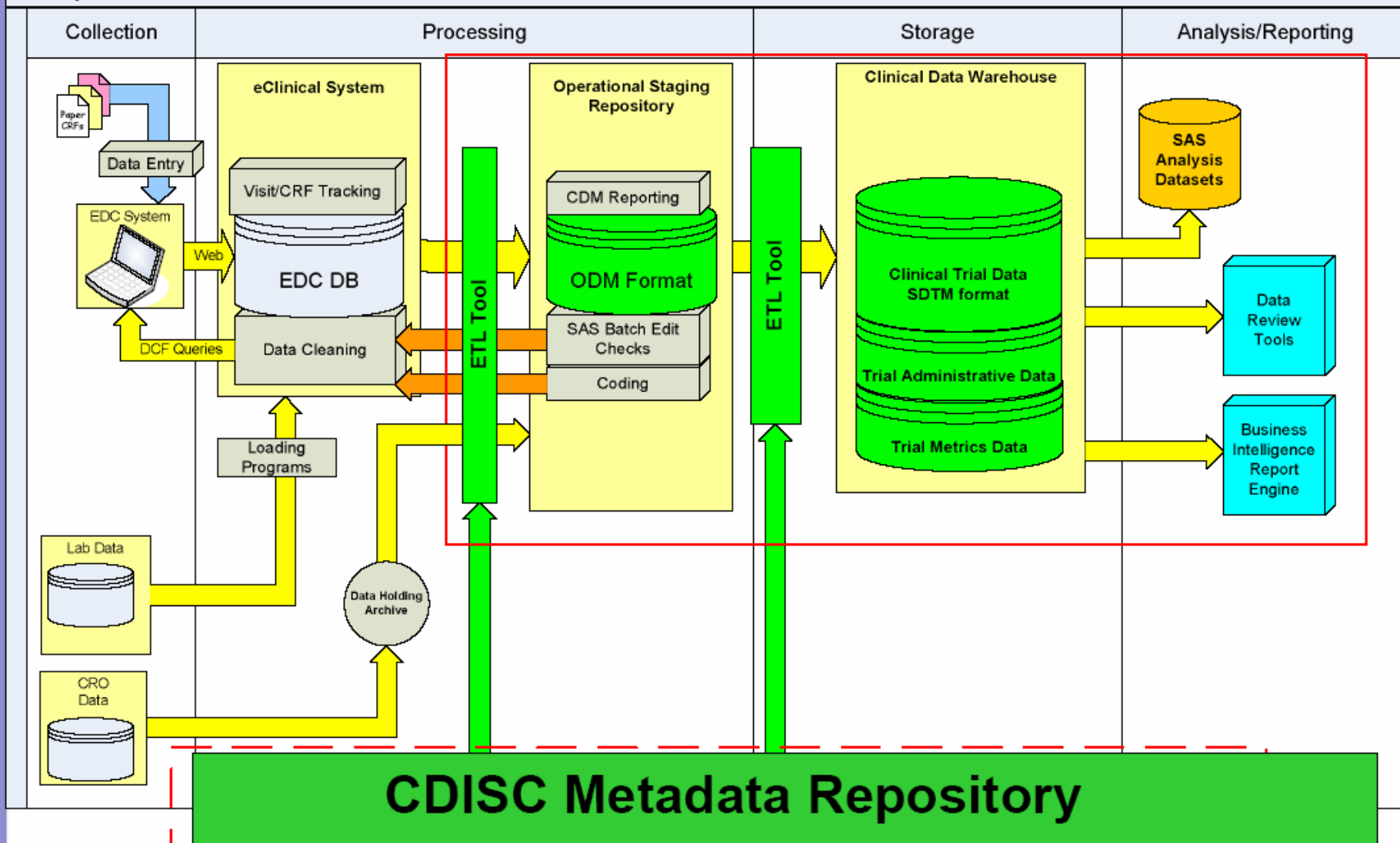
Copyright © 2006



The MIT Information Quality Industry Symposium, 2007



Proposed Clinical Information Process Flow



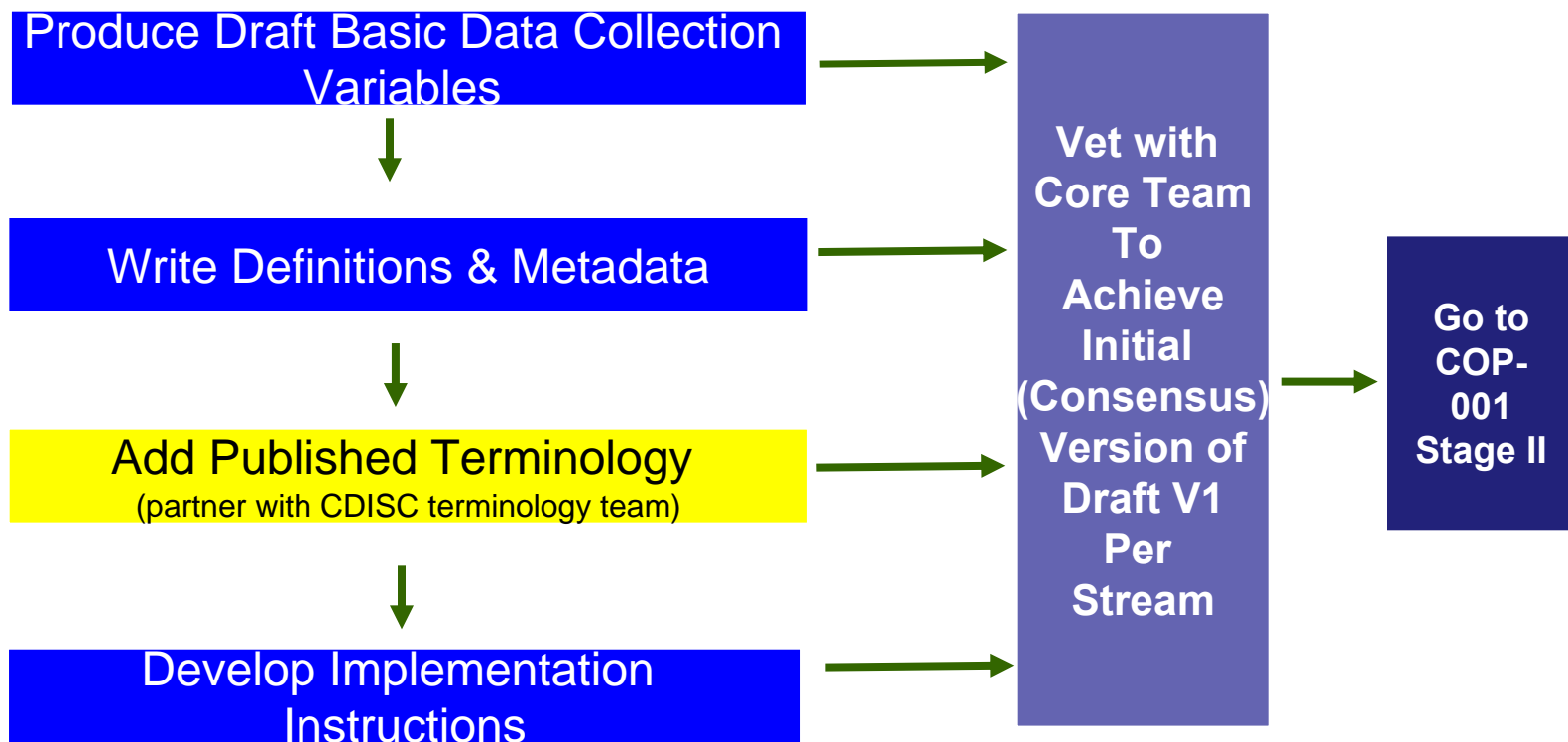
CDISC Interchange 2000-05-27



The MIT Information Quality Industry Symposium, 2007



CDASH Project Goal – To Deliver:





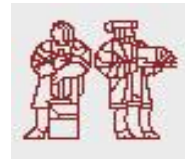
CDASH: CRF standardization guiding principles

- Standard, but flexible
- Comply with regulatory requirements
- Reduce redundancies
- Not duplicate information from other CRFs
- Facilitate collection of meaningful data
- User-friendliness
- Enable easy translation to e-form





The MIT Information Quality Industry Symposium, 2007

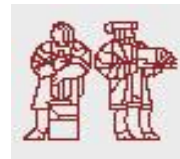


Summary

- Definition of quality
 - “quality data is data that **support conclusions and interpretations equivalent to those derived from error-free data**”
- Steps to make this definition operational
 - Understand sources of errors
 - Identify errors through review/inspection/audit
 - Use inspection results to measure data quality
 - Assess impact of data quality on conclusions drawn from the trial (statistician’s responsibility)
- Challenges: “evidence-based medicine is never obtained” (Dr. J. Woodcock, FDA, 05/10/2007)
- Opportunities: Data standardization. CDISC it!



The MIT Information Quality Industry Symposium, 2007



Thank you!



SOCIETY FOR CLINICAL DATA MANAGEMENT, INC.

Vadim Tantsyura, Head of Data Management

Olive Yuan, Ph.D., Sr. Data Manager

Sergiy Sirichenko, Sr. SAS Programmer

Regeneron Pharmaceuticals

777 Old Saw Mill River Road

Tarrytown, NY 10591

914 345 7626

vadim.tantsyura@regeneron.com

olive.yuan@regeneron.com

sergiy.sirichenko@regeneron.com

