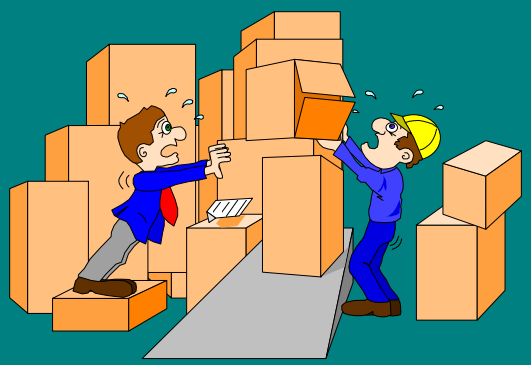
# Accelerating Drug Approvals with Data Standardization: Agenda

- The Case for Submission Data Standards
- The FDA Perspective on Data Standards
- The CDISC Strawman Submission Data Model
- The Process for Defining the Strawman Model
- Panel Discussion

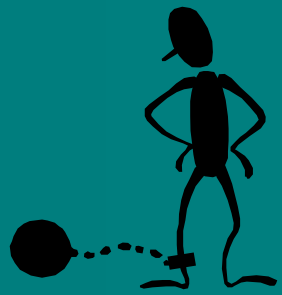
# The Costs of Chaos: The Case for CDM Standards

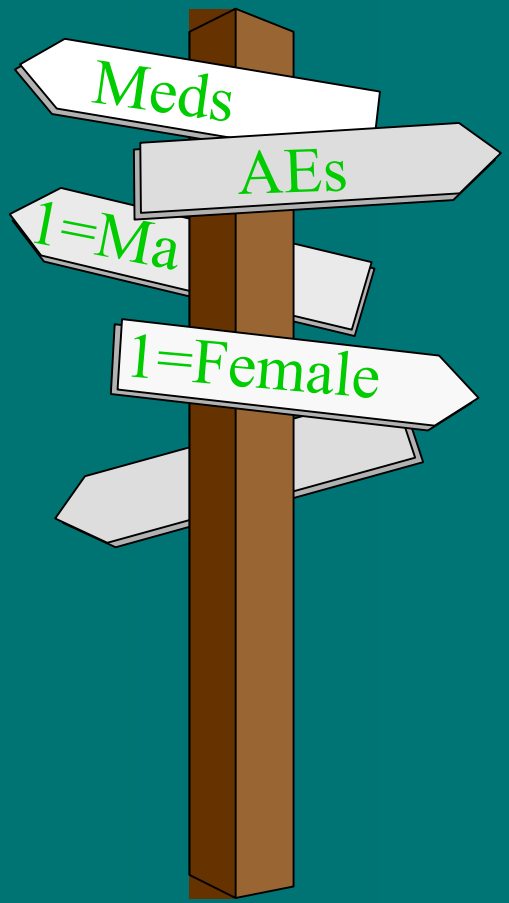


Dave Christiansen, Dr PH
Principal Biostatistician
Genentech, Inc
and
Wayne R. Kubick
Principal
PROsys-LLC

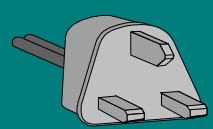
04/08/99

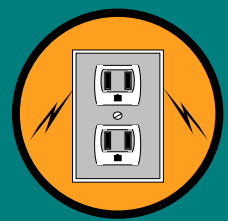
# Life Without Standards






- Requires re-orientation and refitting of analysis tools for each submission review
- Encourages variation within the sponsor environment
- Complicates data integration
- Inhibits major software advances





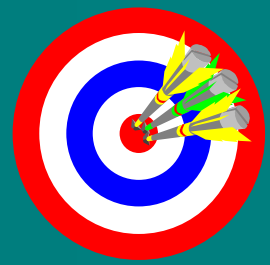
# Previous Responses to Standards Efforts

- Denial -- Too difficult in our Industry
- Solipsism -- The only good standards are our standards -- CANDAs
- Stubbornness -- No one wants to change
- Vendor-phobia -- No one wants to be tied to a vendor
- Regulatory-itis -- Industry resists over-regulation
- Un-Leadership -- Who's willing to go first?
- Low priority -- New Technologies are more interesting than standards.

# CDISC Vision

*To establish standards*
*to improve the process of electronic acquisition*
*and exchange of clinical trials information*

# The Case for Submission Standards

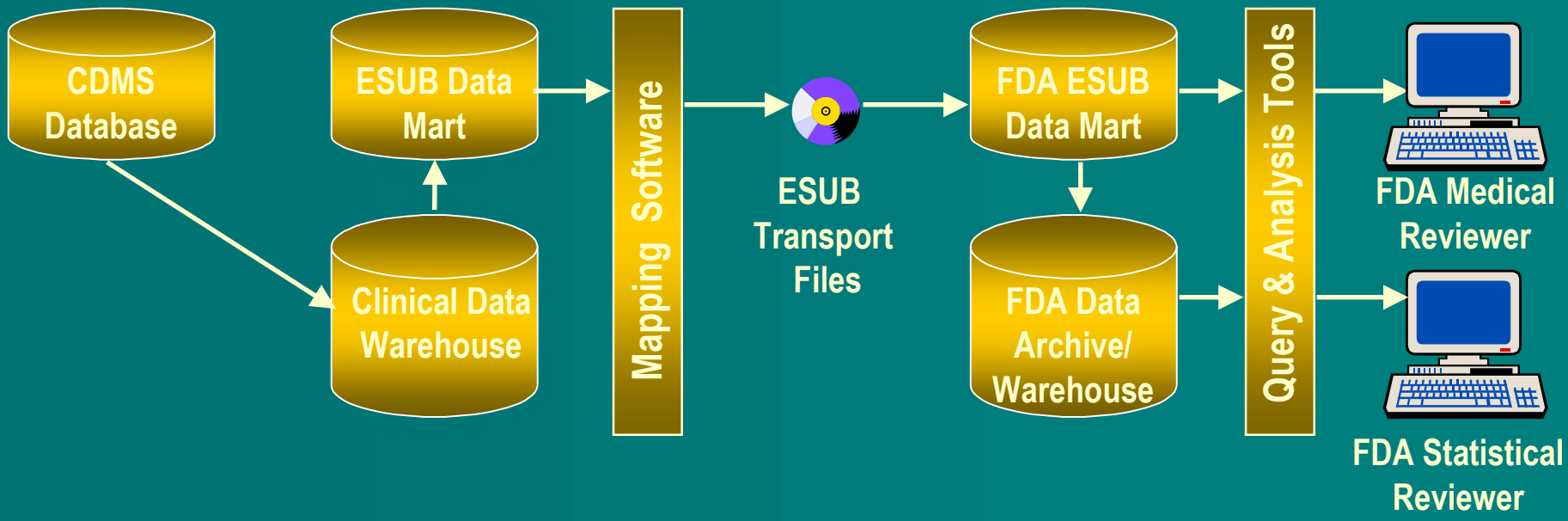


- FDA Guidelines have set the precedent
- FDA is ready now
- Can dramatically improve review throughput
  - Data pre-organized in familiar patterns means less training and fewer misunderstandings
  - Allows FDA to develop their own analytic tools
- Submission interchange standards are attainable - much less complex than a transactional interface
- Gives everyone a common target.

# The Strawman Approach to Submission Standards

- Follow the lead of the FDA Guidelines
- Aim for 80% of domains and 80% of variables
- Define basic metadata standards to guide the organization of common datasets
- Begin developing a library for particular therapeutic areas and aim for a superset of submission standards over time
- Support the FDA's efforts to develop standard access, query and analysis tools based on these standards
- Post standards openly and encourage ongoing input and improvement.

# The Strawman Submission Standards Concept



# The FDA Perspective on Submission Data Standards

- Current Guideline
- Future Desired Direction

03/08/99

# ESUB Data Strawman Approach

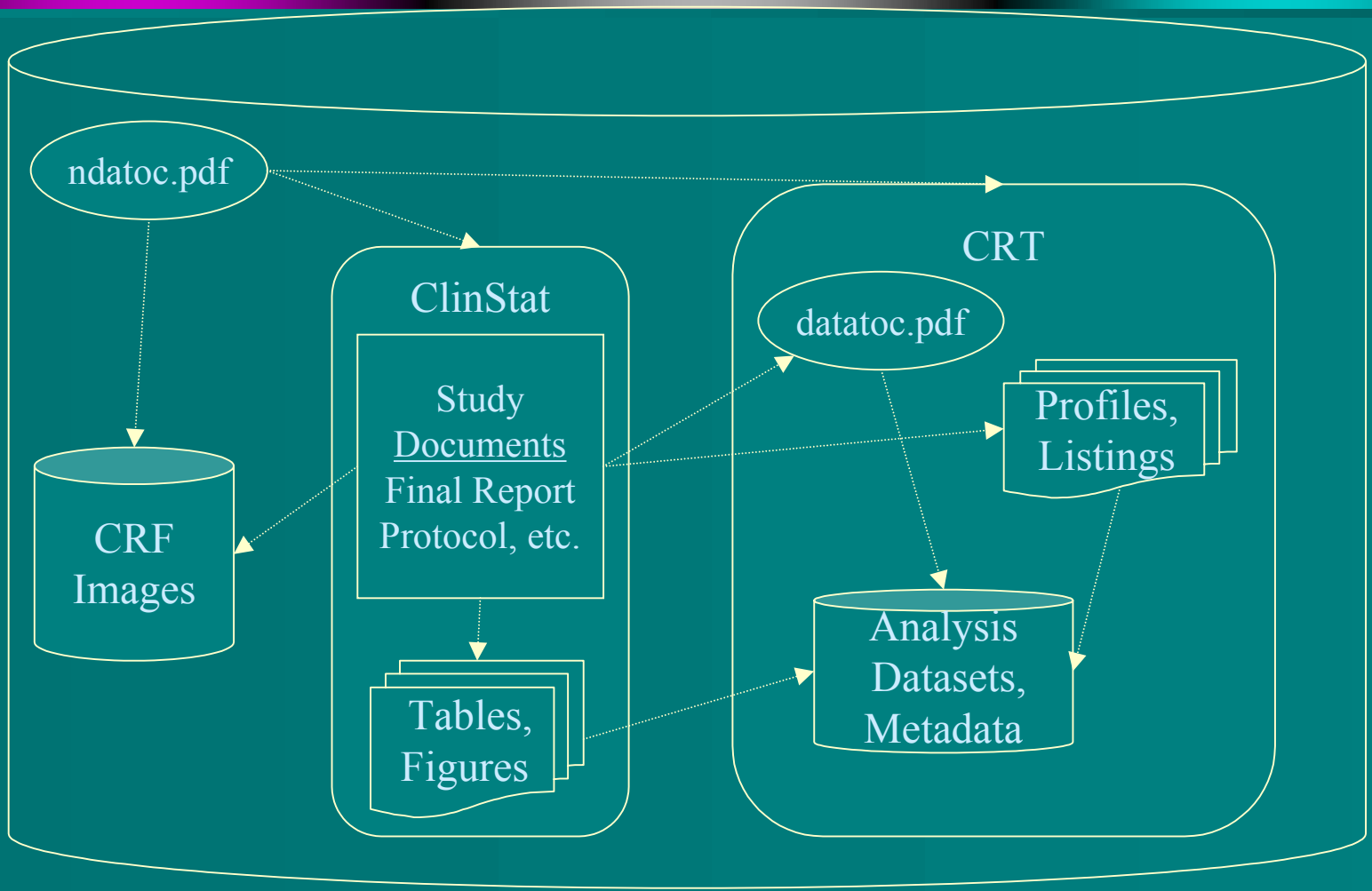
- Organize datasets in folders according to FDA Guidelines
- Define the *Structure* or *Level* of analysis for each dataset
- Classify variables per domain according to *Source*, *Usage* and descriptive *Attributes*
- Link in common *Selection* variables for all datasets
- Use suggested FDA field *Types*, *Codes* and preferred *Labels* wherever possible
- Allow sufficient latitude for adding other variables and domain where scientifically appropriate.

# Taming Chaos: The CDISC "Strawman" MetaData Model

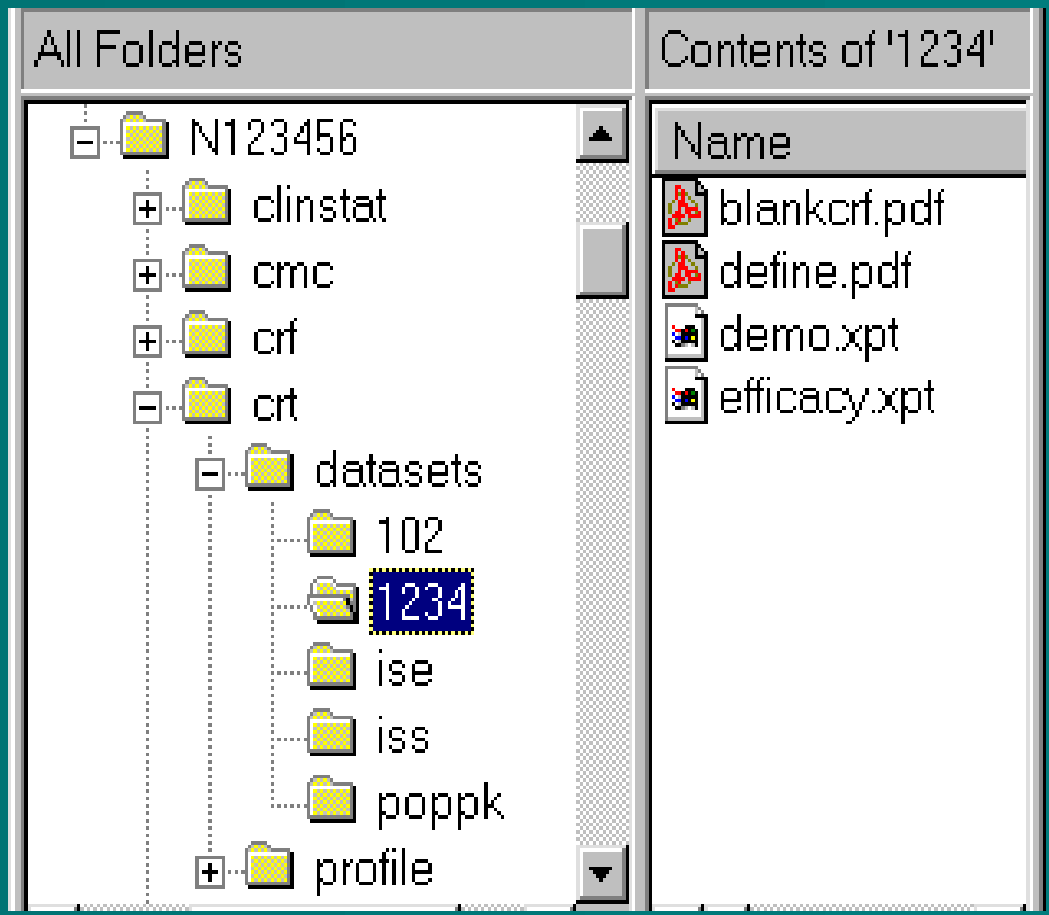
# Data Model Definitions

- **Source Data** - Information collected and recorded about a subject (**Raw data, operational data** or **primitive data**)
- **Derived or Computed Data** - Transformation or reduction of one or more data items by a defined process or algorithm
- **Analysis Database** or **Analysis Files** - A collection of source and derived data items, structured to facilitate data analysis
- **Metadata -** Data about the data; description of the content or purpose of a data base
- **Clinical Trial Data Warehouse -** Analysis files, metadata and documents structured to facilitate the execution and reporting of clinical trials
- **Submission Data Mart -** Subset of a data warehouse specifically designed for submission to a regulatory agency for drug approval

# Submission Data Mart



# Clinical Directory Structure of an Electronic Submission



# Submission Data (and Metadata): Desirable Characteristics

- Should provide clear description of the usage, structure, contents and attributes of all datasets and variables
- Should allow the reviewer to replicate all analyses, tables, graphs and listings with little or no transformation
- Should allow reviewers to easily view and subset the data used to generate any analysis, table, graph or listing

# Metadata: Description and Contents of the Submission Datasets

- Specified in Guidelines
  - Dataset Name (e.g., DEMO)
  - Description (Demographics)
  - Location (crt/datasets/1234/demo.xpt)
- Strawman proposes adding Structure or Level
  - Defines the unit of analysis for a row or observation
  - Useful when multiple datasets are needed for the same clinical domain

# Item Level Lab Dataset: 1 Record/Patient/Visit/Lab Test

| Patient | Visit | Test | Value | Status |
|---|---|---|---|---|
| 1234 | Base | AAA | 95 | Low |
| 1234 | Base | BBB | 122 | Normal |
| 1234 | 1 | AAA | 89 | Low |
| 1234 | 1 | BBB | 153 | High |
| 1234 | 4 | AAA | 91 | Low |
| 1234 | 4 | BBB | 137 | Normal |

# Visit Level Lab Dataset: 1 Record/Patient/Visit

| Patient | Visit | AAA | BBB |
|---|---|---|---|
| 1234 | Base | 95 | 122 |
| 1234 | 1 | 89 | 153 |
| 1234 | 4 | 91 | 137 |

# Dataset Structure Levels

**Patient-Level (1 rec/pat)**

Demographics, Disposition, Inclusion, Exclusion,

**Visit-Level (1 rec/pat/visit)**

Visit Type/Date, Vitals, Efficacy Measurements

**Incident-Level (1 rec/pat/incident)**

AEs, Medications, Diaries, PK, etc.

**Item-Level (1 rec/pat/visit/item)**

Labs, Medical History, etc.

**Other (Look-ups, etc.)**

Subset Patient Lists, Investigator Lists, Cross-Reference Tables...

# Metadata: Variable Description

- Specified in Guidelines
  - Variable Name (e.g., DEMO)
  - Attributes (Label, Type, Codes...)
  - Comments (Source: CRF, derived…)
- Strawman proposes adding Usage
  - Indicates how the variable is used
  - Usage may vary by dataset and analysis

# Data Role Field Classifications

- ***K*** (Key) Variables -- used to uniquely identify and index each record:  Study, Center, Patient ID, Visit, Event Nr
- ***S*** (Selection) Variables -- frequently used to subset, sort or group data for reporting purposes: Sex, Age, Race, Treatment Group...
- ***D*** (Domain) Variables -- variables that relate to the clinical domain and are tabulated and computed for analysis purposes: efficacy measures, lab values, record counts.
- ***D*** (Descriptive) Variables -- provide other reference information or provide input for deriving variables. Help further identify

# Summary of Strawman Data Model

- Standardize **Metadata** content and format
- Add **Structure** to Dataset Description
  - Patient-level
  - Visit-level
  - Incident-level
  - Item-level
- Add **Role** or Classification to Variable Description
  - Key
  - Selection
  - Measurement
  - Reference

# Strawman Definition Process

- Post first cut Strawman for 8-12 domains on DIA web site -- soon
- Collect comments from industry
- Prepare revised version by June DIA/CDISC meeting
- Phase 1: Complete general metadata model and define primary keys for 8-12 critical domains within 6 months
- Phase 2: Continue developing other domains and identify other areas for increased standardization as an ongoing practice.

# Questions for the Panel

- Will the strawman approach work?
- How do we get FDA and industry input?
- Can we really define one set of data that meets both CRT and statistical analysis requirements?
- Can industry really agree on standards for domains, fieldnames?
- Can this be accomplished without inhibiting science?
- Can this be accomplished in our lifetimes?

# Questions and Comments?

Contact:
Wayne-kubick@prosys-llc.com
davec@gene.com
or the CDISC home page:
www.diahome.org/cdisc/dia-siac.htm