

SDTM as an Internal Standard

What's Different?

John Brega
PharmaStat LLC

Presented to the Bay Area CDISC Implementation Network

May 3, 2006

SDTM as an Internal Data Standard?

Yes. SDTM, or a variant of it, will work as an internal standard.

To plan an implementation, however, you need to dig deeper.

- How many *different* applications do you have for a standard? (source data, analysis data, submission data, ...)
- Can a *single* data standard address all your applications?
- For any given application, can you use *conforming* SDTM or does it require *a variant of* SDTM?

Conforming SDTM or a Variant?

For any specific application of a data standard, the choice of conforming SDTM or a variant depends on your answers to these questions:

- What function does the application perform in *your* process?
- What requirements must it satisfy to perform that function?
- Which requirements are addressed by *conforming* SDTM?
- What *nonconforming* changes are needed to address the remaining requirements?

Overview

- SDTM's Intended Purpose
- Some Observations on Observation Classes
- Versions of Study Data in Process
 - Source Data for Analysis
 - Data from Vendors and Partners
 - Data Management Data
 - Analysis Data
 - Submission-ready data
- Conclusions

SDTM's Intended Purpose

- An Electronic Data Interchange standard for submitting data to the FDA
- Designed to support FDA medical reviews
- Intended to represent “fully cooked” data with all the dots connected and ambiguities resolved
- It was *not* designed to function as an internal standard
- Nonetheless, it addresses many requirements

Observations on Observation Classes

- Events (AEs, Medical History, Subject Disposition, ...)
- Interventions (Con Meds, Exposure, ...)
- Findings (Labs, Vitals, PE, Efficacy Results, ...)
- Supplemental Qualifiers

Events Class

- AEs, Medical History, Subject Disposition, ...
- Structure:
 - Study and Patient Identifiers
 - Verbatim term
 - Preferred term and Body System
 - Start and end dates
 - Qualifiers such as Severity, Relationship to Study Drug, ...
- AE data looks just like your old familiar data, with funny variable names.
- “Subject Disposition” data (randomization, end of treatment, termination, death, ...) are all thrown together as events.
Not what you’re used to.

Interventions Class

- Con Meds, Exposure, ...
- Structure:
 - Study and Patient Identifiers
 - Name of treatment
 - Generic name and drug class
 - Start and end dates
 - Qualifiers such as Dose, Indication, Route, ...
- Con Meds data looks just like your old familiar data, with funny variable names.
- Representing raw treatment data may be hard if actual exposure is derived from other values at analysis time.

Findings Class

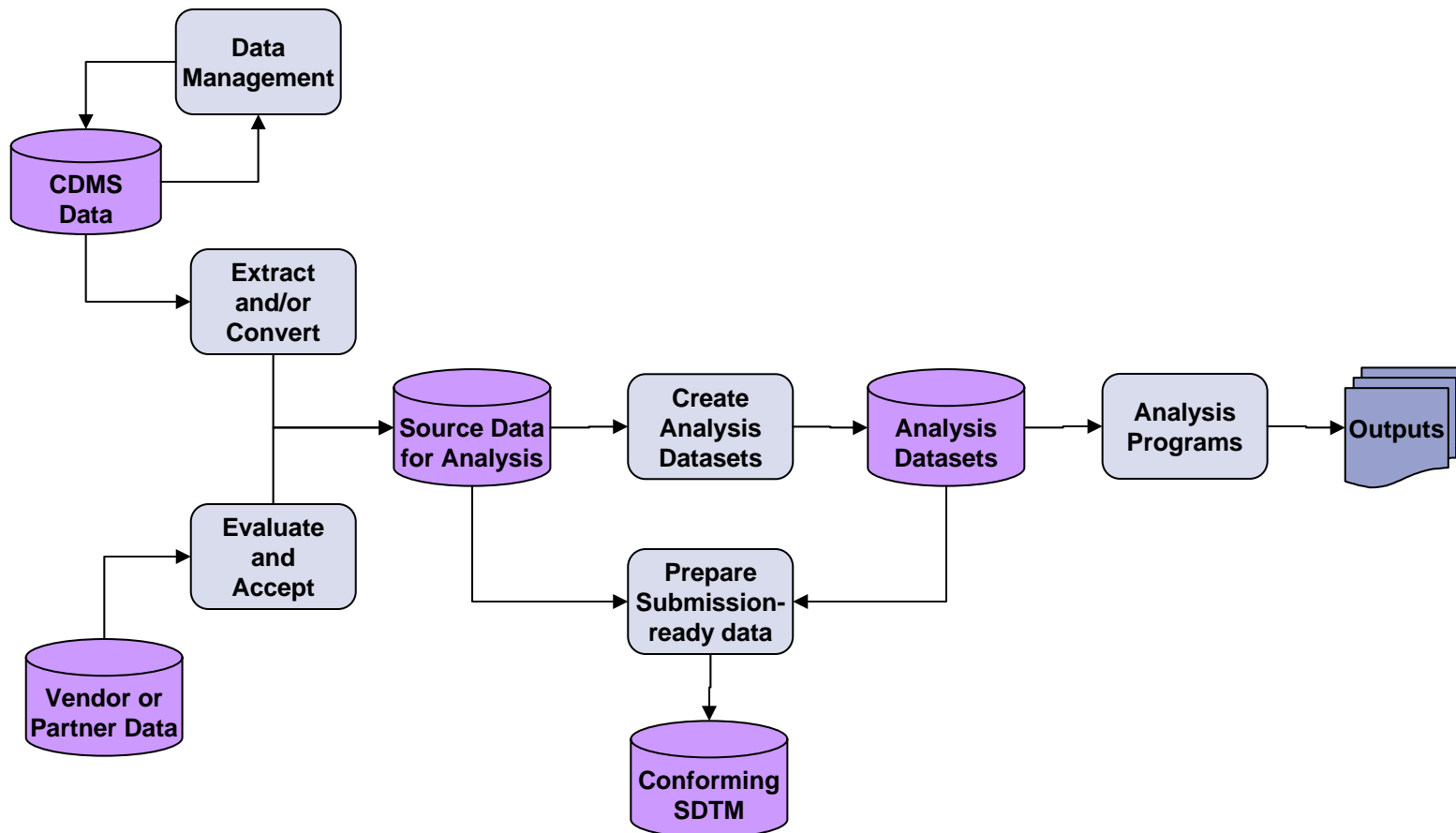
- Labs, Vitals, PE, efficacy measures,...
- Structure:
 - Study and Patient Identifiers
 - Test code and description (e.g., ALKPH, Alkaline Phosphatase)
 - Test value/result (original result and standardized)
 - Date and/or nominal time of collection
 - Qualifiers such as normal range, NR indicator, toxicity grade, ...
- You're familiar with this structure if you've worked with central labs. Works OK for Vitals.
- Ever seen physical exam data presented this way? Weird.
- This structure assumes “tests” are basically independent!

Supplemental Qualifiers

- Additional nonstandard qualifiers (e.g., Other Action Taken)
- Only have meaning when attached to a domain record
- *Must* refer to an existing domain record

Versions of Study Data in Process

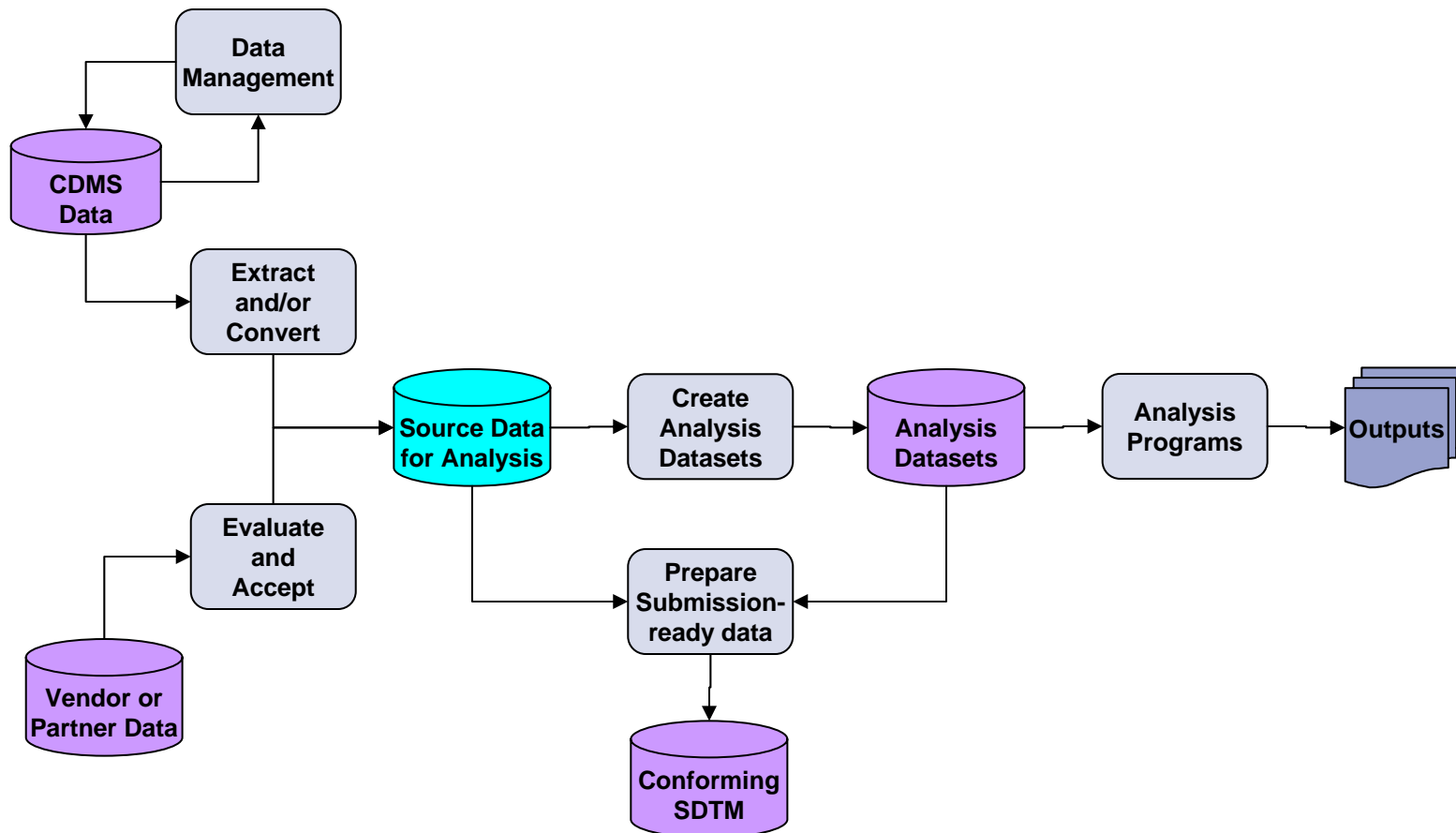
Staging of Data from Collection through Statistical Analysis



Data Versions

- How many versions of a study's data do you have?
- What function does each version perform?
- What requirements must it meet to perform that function?
- Can conforming SDTM satisfy those requirements?
- If not, what are the minimum changes necessary?

Staging of Data: Source Data for Analysis



Source Data for Analysis

- The starting point for analysis
- Usually an extract of the CDMS or a CRO deliverable
- This data should support any analysis permitted by the study design
- What are your requirements?
 - Represent *all* the collected data?
 - Maintain traceability back to the source?
 - Drop data management status variables and codes?
 - Perform only the minimum necessary derivations?

Addressing the Requirements

- When does conforming SDTM *not* handle data you have?
- When does it require data you don't have (yet)?

Data Not Handled

1. The infamous “AENONE” problem
 - A nonconforming problem that requires a nonconforming solution
2. Subjects entering a study more than once
 - You may have to create a third type of Subject ID
3. Verbose inclusion/exclusion criteria and questionnaires
 - Sometimes you just can’t make sense in 40 characters!
4. Raw data from which exposure is derived (this is unusual)
 - The EX domain expects “fully cooked” exposure data
5. Normal ranges separate from lab values
 - Make up your own data model – SDTM has nothing for this

The AENONE Problem

- These are source records which only indicate that no data was collected for the patient
- Example: the “None” box is checked on the AE page
- Is it a *requirement* to represent all the as-collected data?
- If you trust your data management process, do you need a record that tells you there is no record? Why keep it?
- What if you need to do edit checks at analysis time on CRO data that may have unresolved issues?

Possible Solutions

The conforming solution is to drop those records. But...

What if *every* patient has a record indicating no AEs?

- Create a dataset with no observations?
- Have no dataset at all?
- Where do you indicate that this was intentional?

If your requirement is to keep the data, what can you do?

1. Suppqual is not a solution: there's no record to refer to
2. Add an "AENONE" variable to the domain?
3. Put "No AEs Recorded for this Patient" in AETERM?

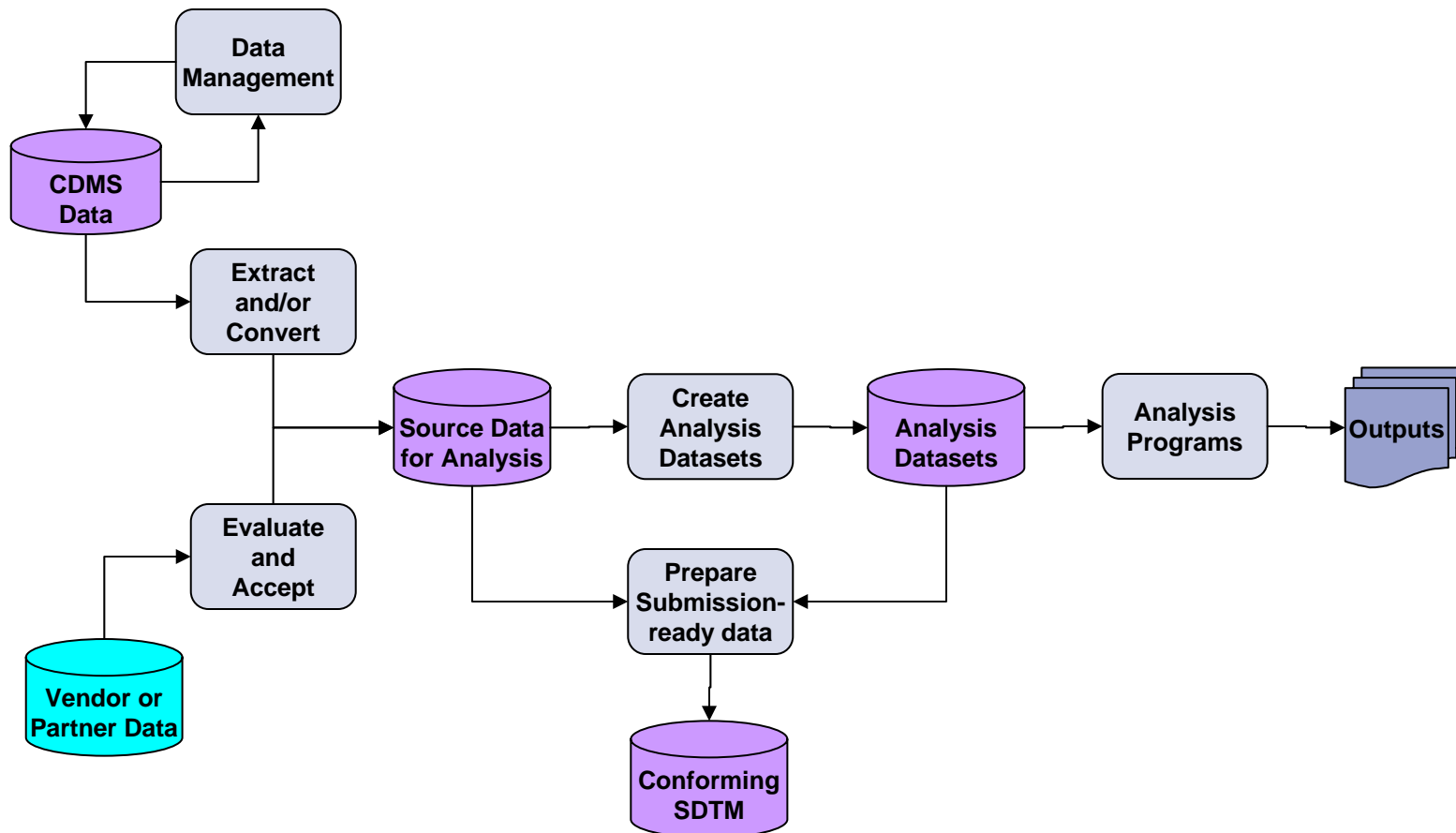
Data You Don't Have (Yet)

- When you create source data for analysis, you may not be able to fill some expected SDTM variables because they are the product of analysis that hasn't yet been done
- Examples include:
 - Dictionary-coded variables for the AE and ConMed domains
 - The “standardized” result values in findings-type domains, including --STRESC, --STRESN, --STRESU, --STNRHI, --STNRLO
 - VISITNUM, VISITDY

Other Considerations

- Will you create and use conforming ISO8601 date/times, or wait until you're preparing data for a submission?
- At what point in your process will you recode variables designated as controlled terminologies?
- For internal purposes you have to control the terminologies of *more* variables than those required by the standard.

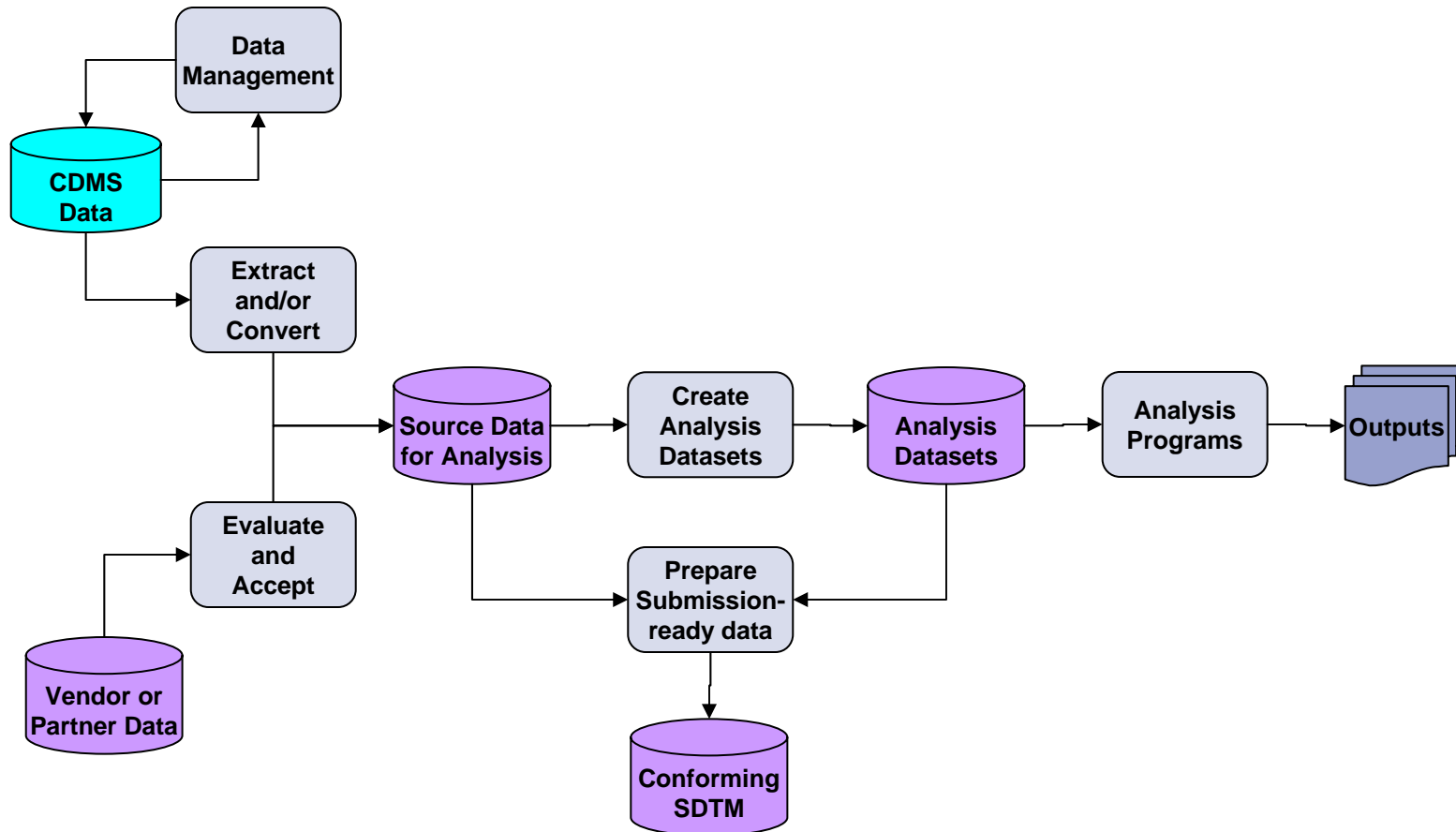
Staging of Data: Vendor or Partner Data



Data from Vendors and Partners

- If you can get whatever you ask for, what should you ask for?
- It depends on how you will use the data...
- What internal uses will it be put to?
 - Analysis?
 - Submission?
 - Archival only?

Staging of Data: Data Management Data



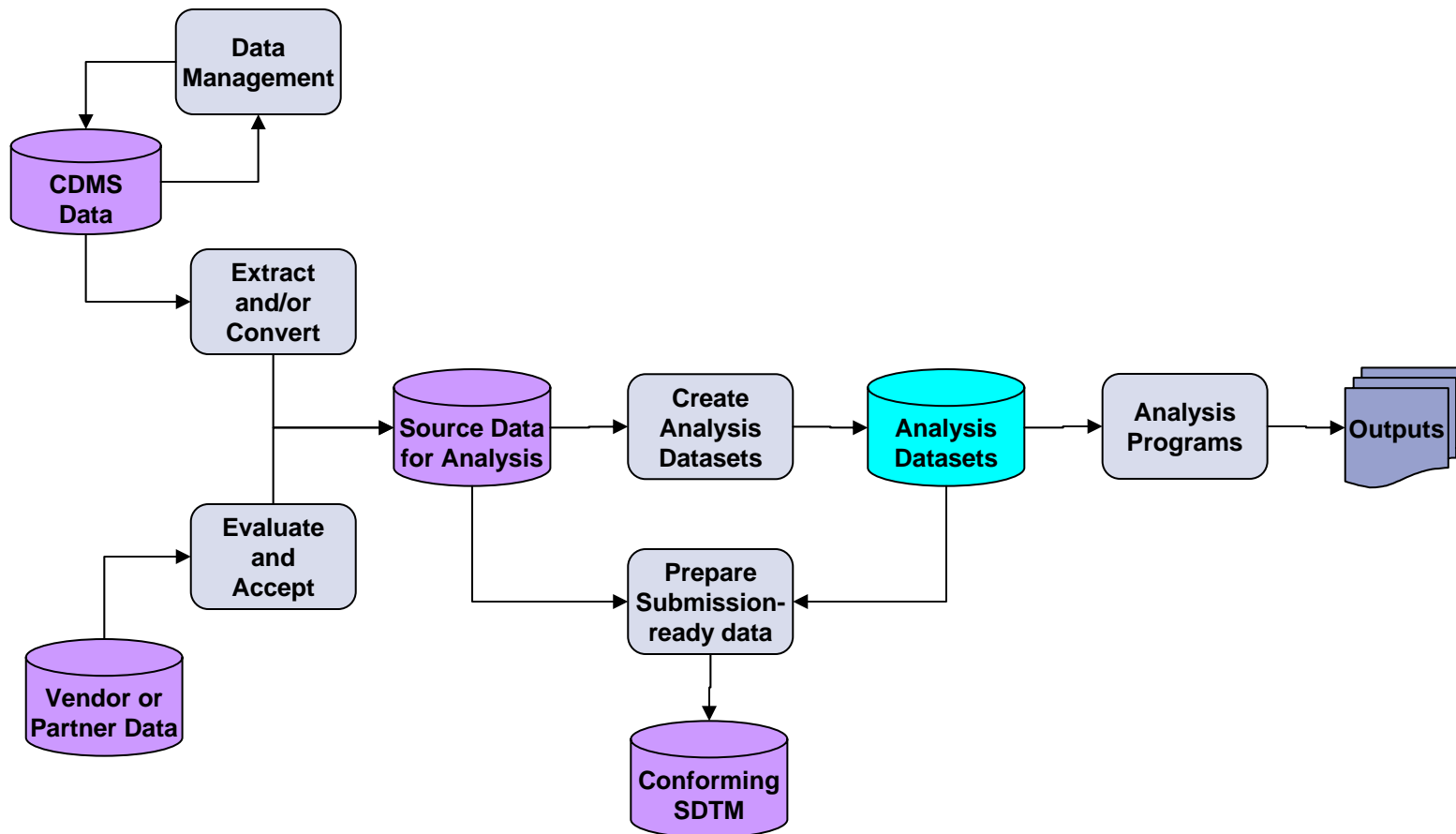
Data Management Data

- The design of this data must support data management functions:
 - Data entry and update
 - Edit checks
 - Review
 - Audit trail
- Data is in-process and incomplete, with few derived variables
- Contains data management process variables (e.g., field status variables) that will not be analyzed or submitted
- A close match with the CRF format may be very important

Data Management Data

- Before pushing SDTM 3.1 structures back into your CDMS data, talk to someone who has already done it and still thinks it was a good idea. Are your circumstances comparable?
- There are several reasonable conventions you can apply in the data management process to set up an easy conversion to SDTM.

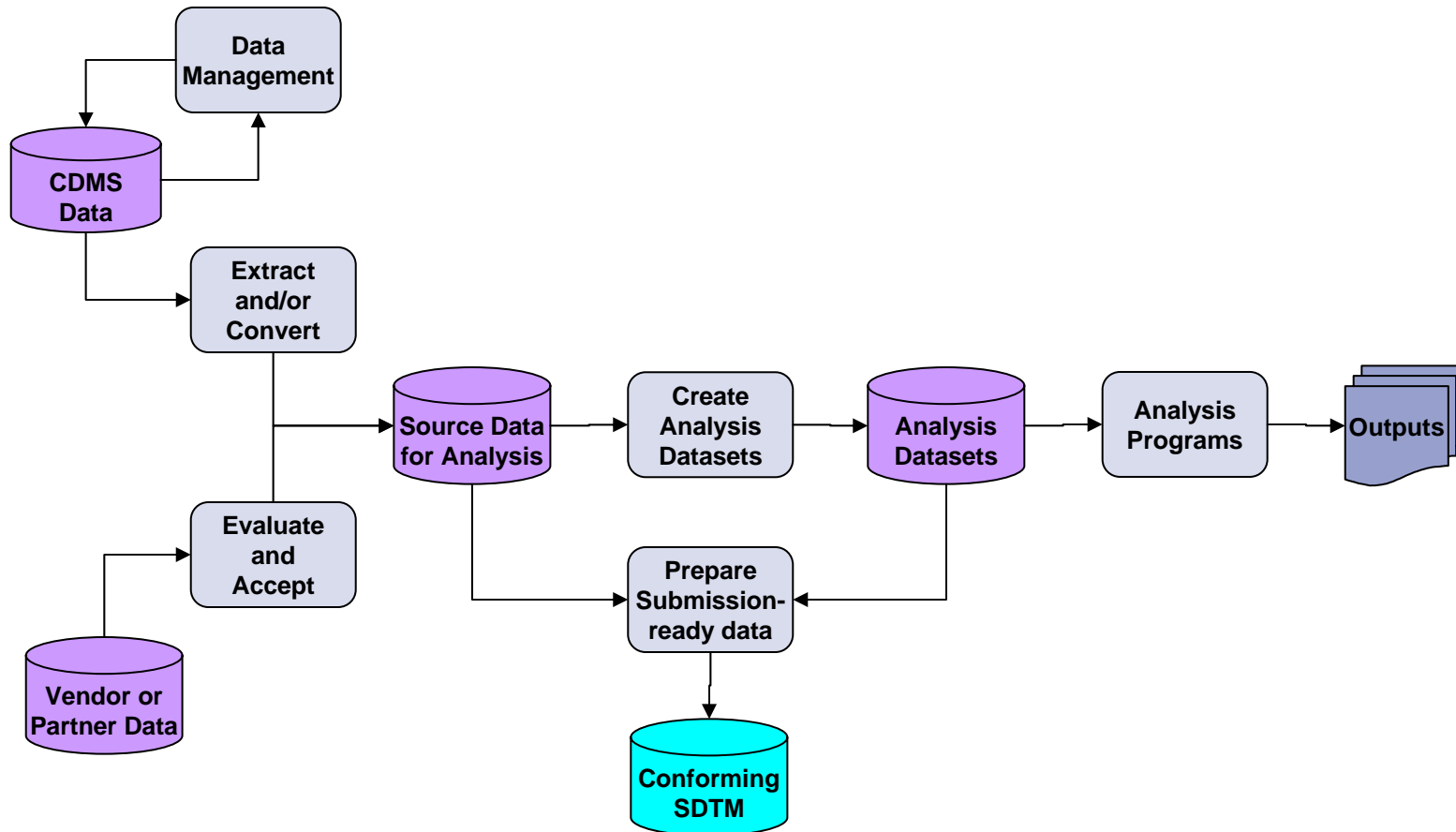
Staging of Data: Analysis Data



Analysis Data

- Analysis requires whole categories of data that SDTM does not account for.
(Change from baseline, subject-level summary statistics, membership in population subsets, ...)
- Requires combining SDTM with ADaM guidelines
- The rules for combining SDTM and ADaM are still being worked out. A CDISC pilot project is underway.
- When the FDA weighs in things may change, as they did from SDS 2.0 to SDTM 1.0 (SDTM IG 3.0)
- Current ADaM guidelines address internal uses. An FDA-mandated standard will be a data interchange standard.

Staging of Data: Submission-ready Data



Submission-ready Data

- No ifs, ands or buts, this will require fully conforming SDTM when the FDA completes its rule-making process and publishes the regulation, perhaps next year.

Conclusions

- There are places for both conforming SDTM and SDTM-based data standards in internal operations.
- Use of strong standards yields process clarity and efficiencies far beyond the obvious ones. The benefits are powerful and pervasive.
- When appropriate, the closer you get to strict conformance the greater the benefit .
- The issues raised here are *not* the tip of the iceberg. They reflect experience on over 150 studies from more than 20 organizations.

Contact Information

John Brega

PharmaStat LLC

JBrega@PharmaStat.com

www.pharmastat.com

510-482-2432

