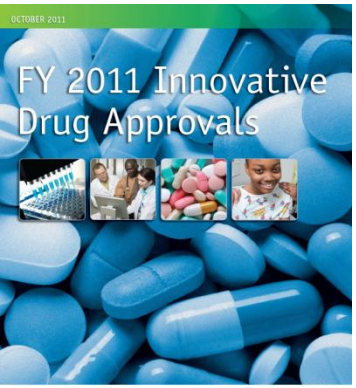


Cost-Benefit Analysis of Retrospective vs. Prospective Data Standardization

Vicki Seyfert-Margolis, PhD
Senior Advisor, Science Innovation and Policy
Food and Drug Administration

IOM Sharing Clinical Research Data
Oct 4, 2012

1

Sponsor

Tables & Listings (flat files)

SDTM V1.1,1.2,3.1.1

2

Reviewers

Tables & Listings (flat files)

SDTM V1.1,1.2,3.1.1

3

Sponsor

Information Exchange using HL-7 V3.0 cDISC Data "RAW"

4

Computational Biology (R&D) Cross-Trial Comparisons

5

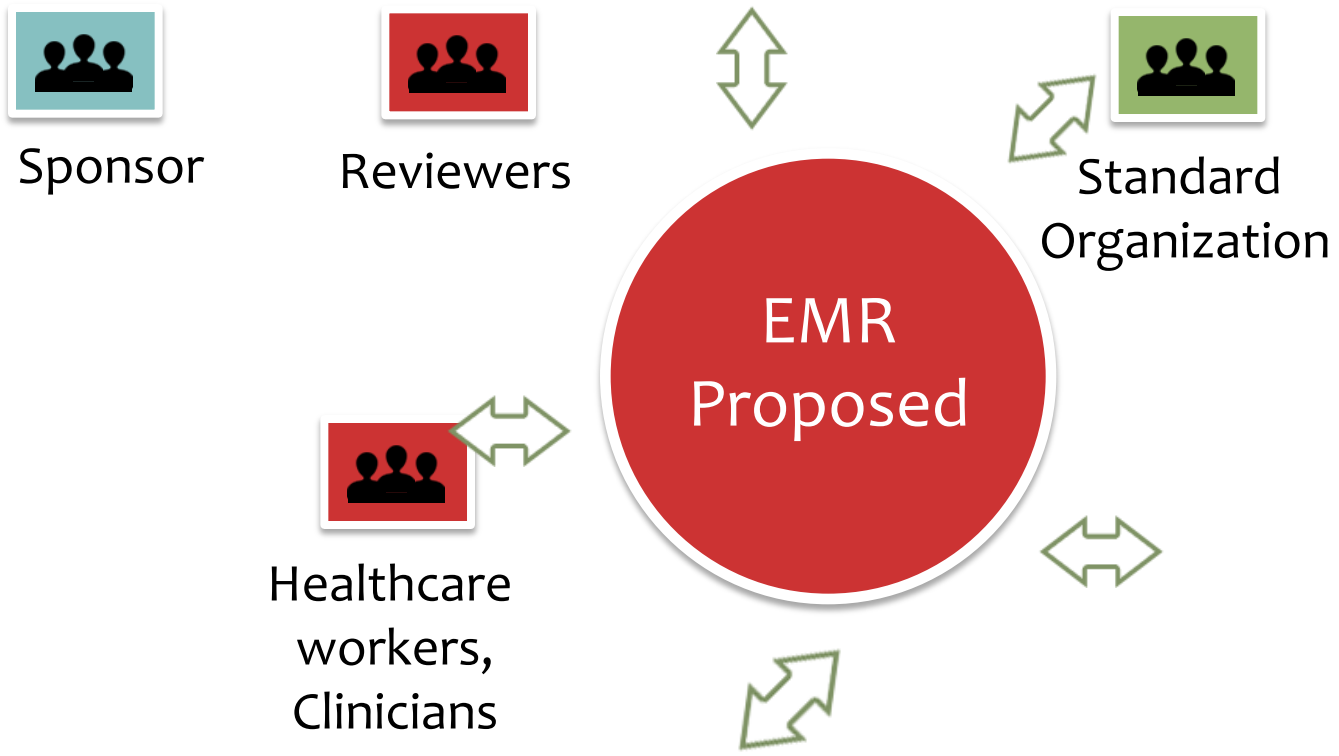
Reviewers

SAS Programming at FDA

HL-7 v2.0 Hospitals, Clinics

EMR PROPOSED

-Access to Analysis tools to manipulate data
-Query FDA for Data Views



5

Regulatory Agencies

We tried several approaches:

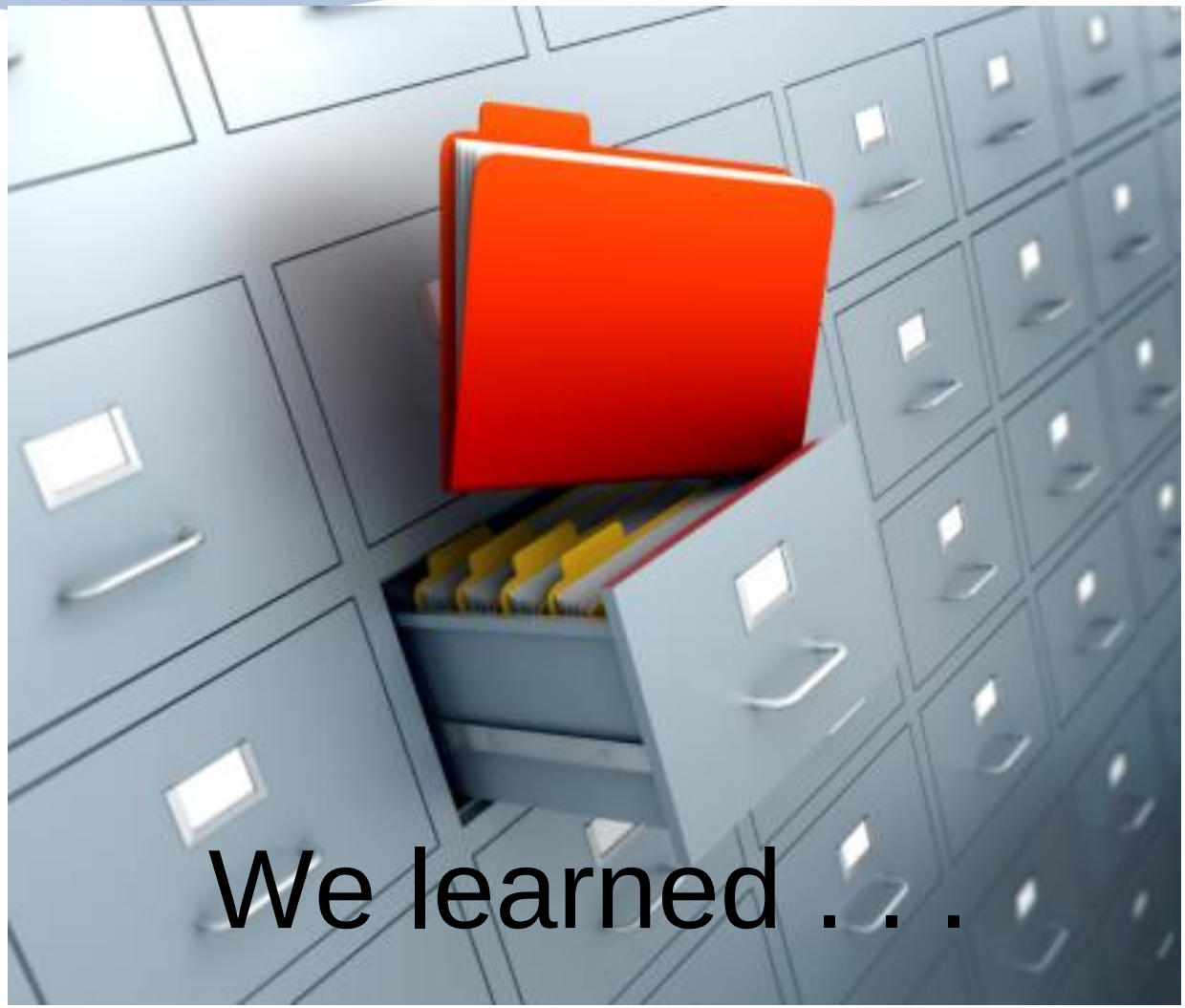
- 1)Legacy Data Conversion – convert all data to a standard format (without no predetermined scientific question)

- 2)Amalga™ – use converted data and/or unconverted data to answer a specific scientific question

Approach #1

Legacy Data Conversion Project -

- Support the conversion of 101 legacy clinical trial data to the SDTM and ADaM formats to enable exploration of PCOR questions related to vaccines, drugs, and medical devices.
- Provide FDA data to support analysis across studies, products, and therapeutic areas.
- It is important to understand the difference between this conversion activity (LDC to support CER) and a sponsor's activities in support of regulatory submissions


We learned . . .

Resources

- Scientific questions drive details in the conversion
  - Clinical/Scientific expertise required to determine how to re-organize the data to fit in the standard
- Terminology/Dictionary harmonization requires clinical expertise
- Statisticians required to translate questions into analyzable components
- QC of converted data is essential but time-consuming.
- Conversion activity is resource intensive and expensive.

Data Quality

- Data quality and harmonization are fundamental to successful data analysis.
- Quality of standardized data might best be achieved during planning and collection steps.
- If data standards are not considered before and during collection, truly standardized data may not be possible.
- Standardization and quality of data are not synonyms
  - Standardization doesn't ensure quality
  - If not done well, conversion to a standard format has potential to adversely affect data quality and analysis.

Fit for Purpose

- Standardization does not imply that data is "fit for purpose".
  - Standardized data may or may not answer our PCOR questions.
  - Data standardized for PCOR may not be useful for future analysis.
  - Can converted data be so fit for a specific purpose that it is not otherwise useful?
- In some instances, conversion to a standard (especially when converting data for a specific goal or purpose) may result in a loss of traceability from the source data or CRF

Collection, Submission, and Conversion of Standard Data

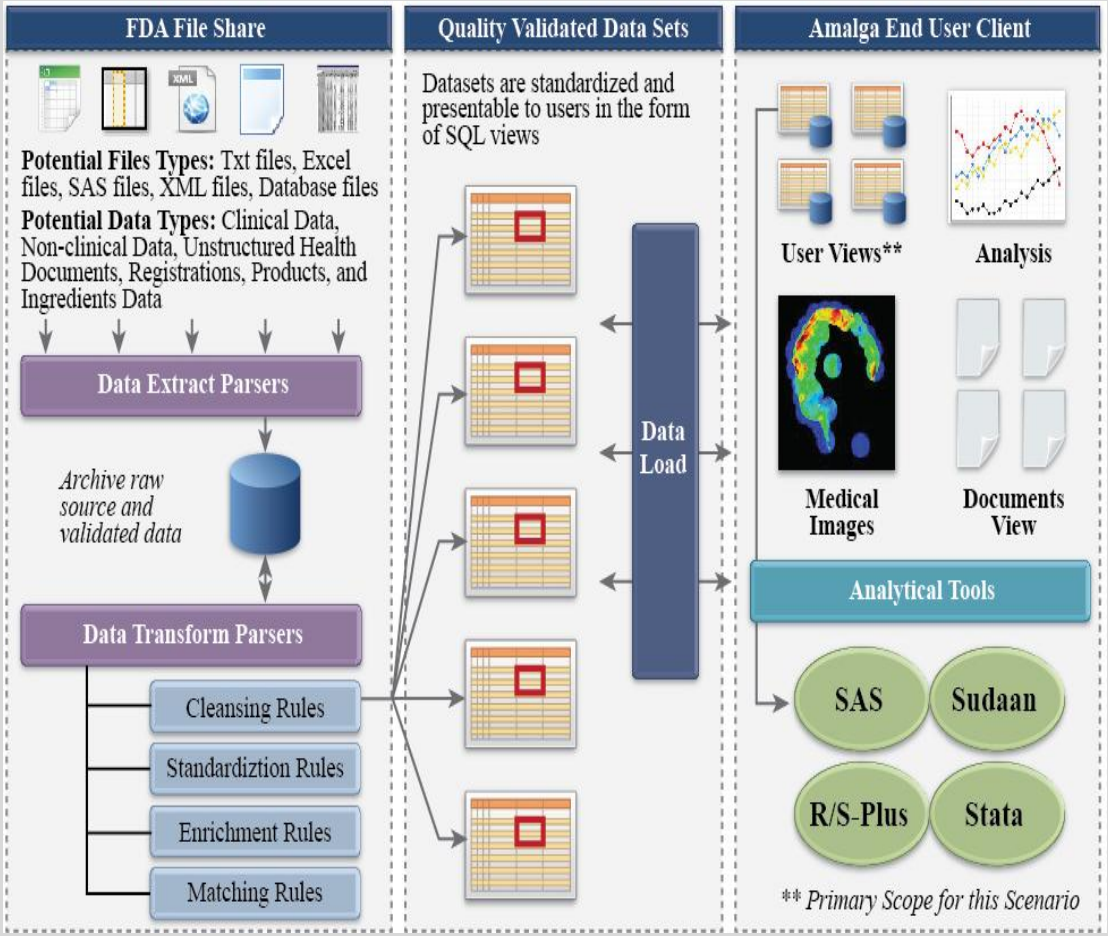
- Not all collected data must be submitted
  - Ideally, initial study planning phases could exclude data that FDA does not need or want
  - FDA does not want subject initials
  - In some instances, original data was unnecessarily confusing, so not converted.
    - For example, the original term "gypsy" was converted to "unknown" in race field
- Not all data must be standardized
  - FDA is working to identify minimum set of data points that must be standard for analysis.

Parting Thoughts

- Collection of data using standard (vs. conversion to a standard) is optimal.
- The standard should be implemented in the same way across studies.
  - Centers created business rules
    - FDA business rules could help Sponsors use the existing standards in a way that facilitates analysis across studies at FDA.
  - Might be specific to centers or therapeutic areas
- Standardization allows for the identification of areas for improvement in the clinical data lifecycle
- Standardization will not solve all problems with study data, it does however illuminate many of them
- Illumination lights the way for improvements

Approach #2

Amalga™ as a platform enables end-users to view data from disparate sources, generate dashboards and reports, and export the corresponding data using the web interface



Amalga™

- Amalga™ is a commercial off-the shelf (COTS) product that leverages customized, built-in message parsers to ingest disparate data sources and images, generates dashboards and reports of the reconciled information, and exports the corresponding material using a web interface
- Historically, utilized in Electronic Health Records (EHR) and in-patient care settings, this was the first instance where Amalga™ was implemented in a regulatory environment
- The one year pilot, which culminated in the successful integration and analysis of disparate regulatory data sets encompassing the medical product lifecycle, demonstrated the feasibility of leveraging Amalga™ to enhance the quality, efficiency, and accuracy of FDA reviews

Non-standardized data structures hinder CDRH in processing regulatory data submissions

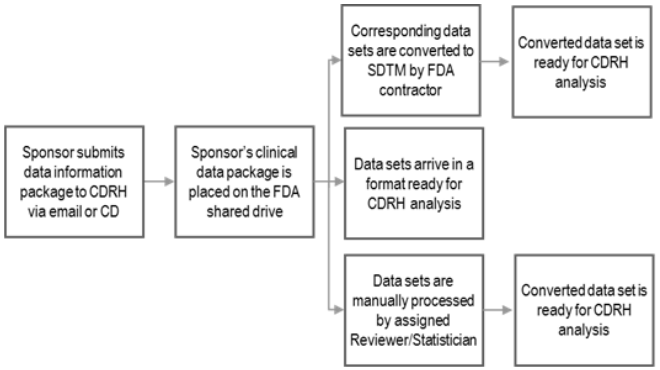
Regulatory Data Packages

- Sponsors typically submit their clinical and post-marketing data packages to CDRH via email or CD in multiple file formats including:
  - SAS (as SAS datasets or SAS XPORT files)
  - Spreadsheets (Microsoft Excel or other)
  - S-Plus or R files
  - XML files
  - ASCII flat files (comma or tab-delimited)
- Lack of standardized data and file formats impede the Agency's review process by adding extraneous steps and time
- CDRH employs contractors to convert the large and disparate data sets submitted in multiple file formats into SDTM for representation of clinical trial tabulation data

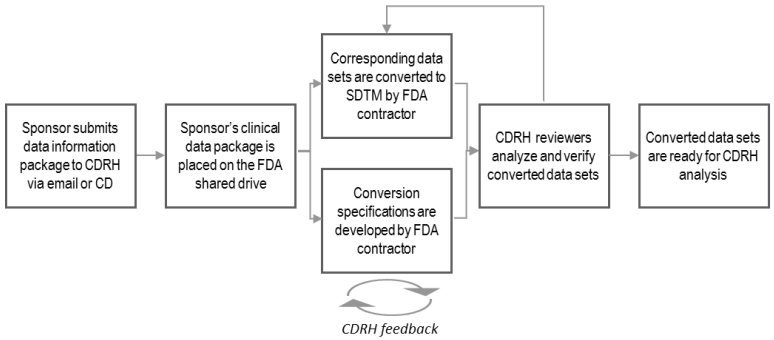
Post-Marketing Safety Surveillance Reports

- CDRH post-marketing safety evaluators have had difficulty efficiently analyzing thousands of safety reports generated by the MAUDE Database and linking specific observations to clinical data sets
- Extensive time and resources are needed to integrate, reconcile, and analyze information between narrative-laden MAUDE reports and existing clinical data sets for marketed device products

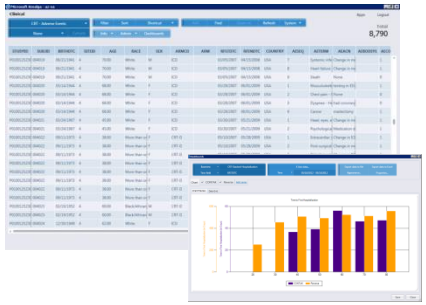
Regulatory Data Conversion Process

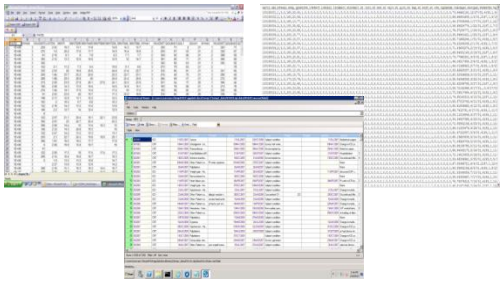


Contractor Legacy Data Conversion Process



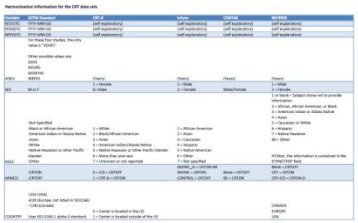
The technical approach consists of five steps that integrates disparate file formats and sources into user-friendly views



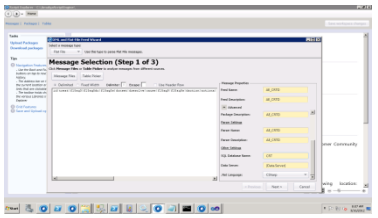


1. Obtain Clinical and Post-Marketing Data Sets

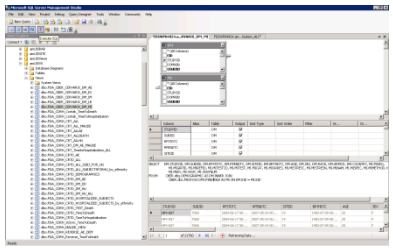
2. Develop SDTM Mapping Schema and Harmonize Data Sets



3. Develop Clinical and Post-Marketing Data Set Parser



4. Ingest Data Sets into Amalga™ and Create SQL Database Views



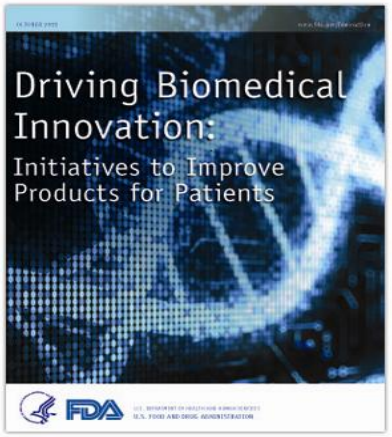
5. Generate Amalga™ User Views, Dashboards, and Reports

Microsoft® Amalga

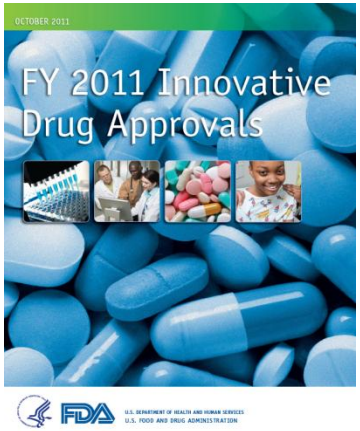
We have identified a set of common challenges and potential integration points across our scientific computing efforts…

	Challenge	*Sample Effort*
Data Integration	Inability to integrate structured and unstructured data due to lack of standards, frameworks, and technologies	Develop an event-centric ontology to represent severe septic shock in healthcare delivery settings
Compute and Store	Insufficient capacity to host and process “big data” to address research and regulatory questions	Develop large database analysis solution using map and reduce jobs on the CDRH computing cluster
Analytical Methods	Gaps in analytical techniques and models to confirm or deny safety, effectiveness, and compliance	Implement seamless text mining and NLP solution to enable text extraction from unstructured VAERs fields
Selecting Problems to Solve	How to identify appropriate inductive and deductive questions to address in analysis of large scientific computing efforts	Identify opportunities to automate regulatory processes and scientific research with standardized data sets


Advancing Regulatory Science at FDA


Driving Biomedical Innovation:
Initiatives to Improve Products for Patients


FDA
Regulatory Science in FDA's Center for Devices and Radiological Health:


OCTOBER 2011
FY 2011 Innovative Drug Approvals