

CDER, Regulatory Science Informatics, and Metadata: What We've Learned

Salvatore Pepe, PharmD, MS

FDA, CDER, Office of Translational Sciences,
Knowledge Management Team

June 2, 2016

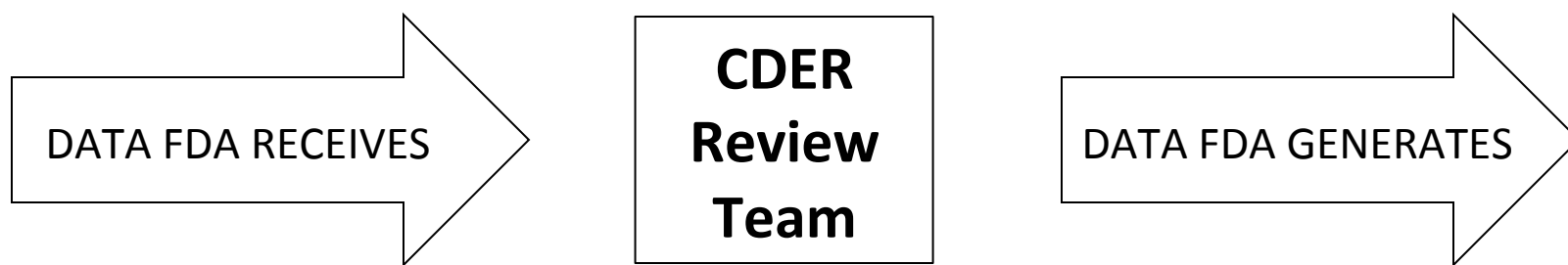
Disclaimer

The opinions expressed in this presentation are solely mine and do not represent the viewpoints of the FDA

Background

- There are many components of the drug development process and lifecycle
- FDA CDER's new drug review process is extremely complex, with many streams of data from multiple sources
- When attempting to use data to understand and characterize the review process and leverage data for knowledge management, it is imperative that these data streams are also understood and well-characterized

Two Major Data Streams



- Patient level datasets
- Administrative documents
- Adverse event reports
- Safety reports
- Citizen's Petitions
- Study protocols
- Meeting requests
- Applications (IND/NDA/BLA/ Supplemental)

- Discipline reviews
- Administrative documents
- Protocol assessments
- Legal correspondence
- Information requests
- Meeting minutes
- Other internal documents

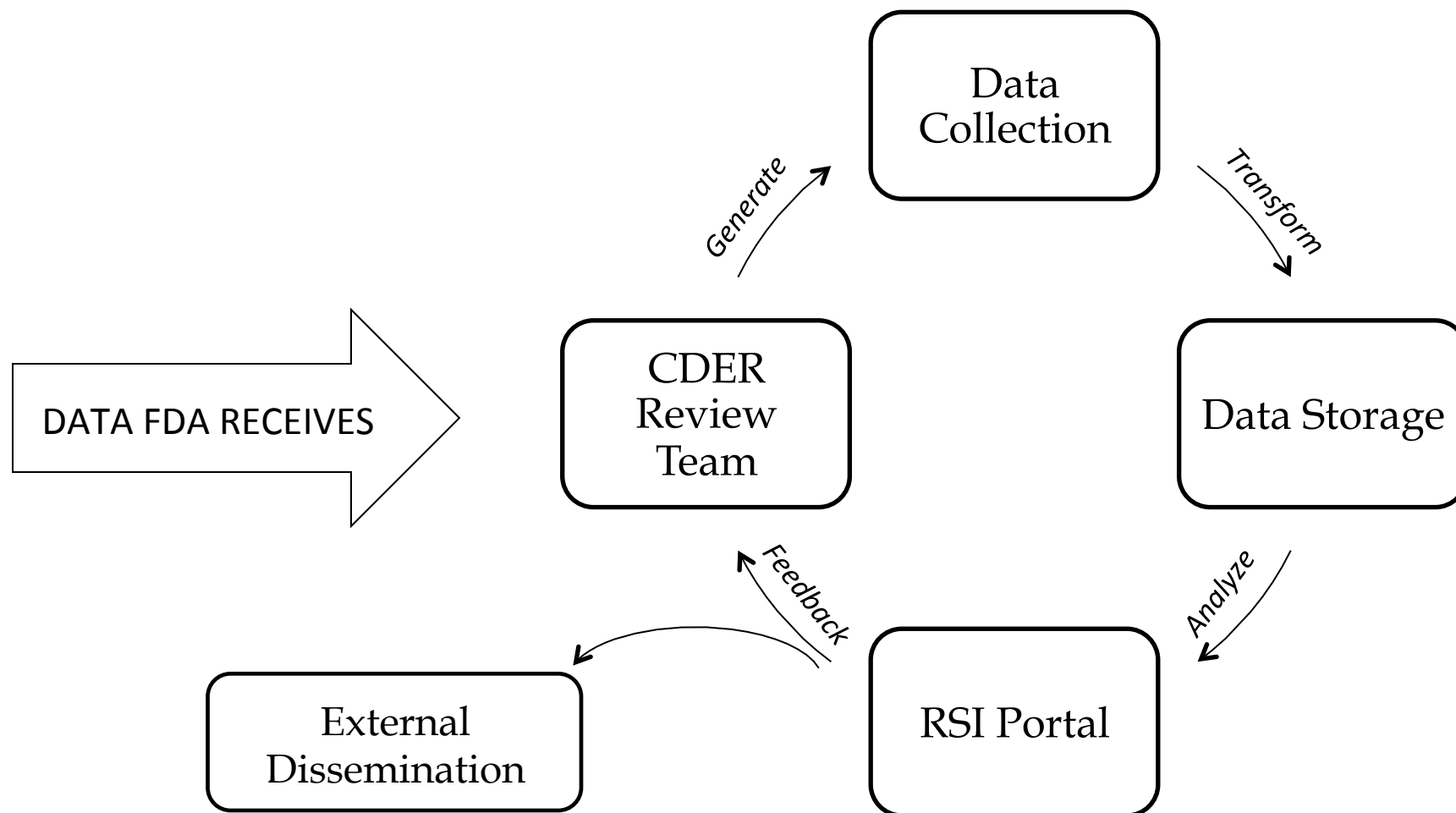
Challenge

- Much of the data standardization efforts focus on data that is submitted to CDER
 - Helps facilitate efficient and standardized regulatory review of new drugs and biological products submitted by many external stakeholders
- Regulatory science information (RSI) that CDER generates is a less mature area of informatics
 - Generally exists in archived free-text documents
 - Information tends to be unstructured
 - Few existing standards to characterize internal data
 - Efforts to extract and analyze information tend to be manual and retrospective in nature

Opportunities

- Many standardized document templates currently in use
- Programmatic information collection efforts upon which to build
- Staff engagement is at critical mass
- CDER-generated information can inform:
 - Congressional requests
 - Press inquiries
 - Mandatory legislative reporting
 - Inform policy and legislative negotiations
 - Regulatory science research

Our Vision



Our Initial Approach

What Questions Do We Need To Answer?



What Data Are Needed To Answer Those Questions?



What Is the Primary Source of That Data?



How Do We Capture/Store That Data?



How Do We Make That Data Available To Those That Need It?

The DASH Database

- What is DASH?
 - A regulatory science database that contains marketing application and supporting drug development information derived from CDER review and regulatory documentation*
- Why was DASH created?
 - Initially, as part of FDASIA for the Rare Diseases Program to track rare disease drug applications
 - Later, to help CDER understand and characterize new drug and biological product review

** Contains commercial confidential information, for internal use only*

What Data Are In DASH?

- Scope
 - New Molecular Entity (NME) New Drug Applications (NDA's) and Biologics Licensing Applications (BLA's) since 10/1/2007
 - Efficacy Supplements to NDA's/BLA's, limited to New/Modified indication or significant change to patient population (FYs 2012-2014)
- Data Buckets
 - Investigational New Drug (IND) Application Information
 - NDA/BLA Information
 - Drug Properties
 - Sponsor Information
 - Clinical Trial Information
 - Other Miscellaneous/Application Level Information

Novel Data Management Approach

- Defined data elements
- Incorporated standardized terminologies for identified data elements and master data management
 - Adopt existing standards (eg., available nomenclature)
 - De-novo standards creation
- Crafted a comprehensive SOP detailing:
 - Scope of the database
 - Processes and procedures
 - Data elements, primary source, collection methods
 - Business justification
- Formed multi-disciplinary committee to adjudicate and curate subjective data elements

What We've Learned

What Questions Do We Need To Answer?



What Data Are Needed To Answer Those Questions?



What Is the Primary Source of That Data?



How Do We Capture That Data?



How Could We Capture That Data Efficiently?



How Do We Make That Data Available To Those That Need It?

What is “Efficiently?”

- Capture data in a standardized fashion...
- Identify individual, siloed data gathering efforts...
- Collaborate with other data gathering groups...
- Utilize technology in place of manual resources...
- Capture more information with existing resources...
- Have subject matter experts collect within their expertise...
- Capture data at the point of generation...
- Work prospectively instead of retrospectively...
- Avoid double data entry...
- Feed information, and thus knowledge, back into the system...

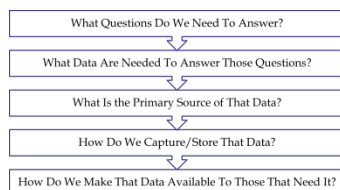
First Step...Metadata

- Critical component of understanding and characterizing CDER's RSI
- When scaling up to CDER-wide, metadata efforts and scope expand tremendously
 - Collaborate with existing metadata stakeholders in CDER
 - Identify best practices that serve all
 - Collect CDER's metadata using a centralized metadata approach

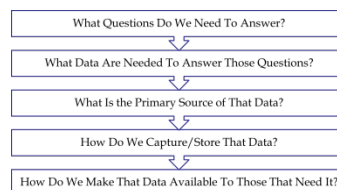
Centralized Metadata Approach

Early Stages - Find, Inventory, and Harmonize Existing Metadata

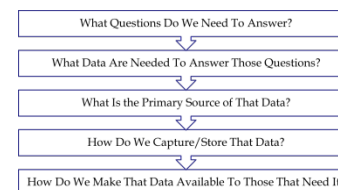
PROJECT A



PROJECT B



PROJECT C



METADATA REPOSITORY



What We Hope To Do...

- Leverage information from existing policies and data standards wherever possible
- Identify high-value metadata elements
- Eliminate redundancy and duplicative efforts by having a centralized metadata repository harnessing collaborative efforts
- Identify opportunities for workgroups, data stewards, and governance bodies
- Explore automated, source-level data capture
- Create seamless data management principles for retrospective (legacy) and prospective data collection

Conclusions

- Gathering metadata is a critical component of understanding and characterizing CDER's data streams
- DASH could serve as a model to inform future data collection and knowledge management efforts
- In order for a large-scale RSI effort to be successful, we need to incorporate efficient data management principles

Thank You!

Questions?

