Commentary

Regulatory Administrative Databases in FDA's Center for Biologics Evaluation and Research: Convergence Toward a Unified Database

Jeffrey K. Smith^{1,2}

Received 16 August 2012; accepted 10 December 2012; published online 27 December 2012

Abstract. Regulatory administrative database systems within the Food and Drug Administration's (FDA) Center for Biologics Evaluation and Research (CBER) are essential to supporting its core mission, as a regulatory agency. Such systems are used within FDA to manage information and processes surrounding the processing, review, and tracking of investigational and marketed product submissions. This is an area of increasing interest in the pharmaceutical industry and has been a topic at trade association conferences (Buckley 2012). Such databases in CBER are complex, not for the type or relevance of the data to any particular scientific discipline but because of the variety of regulatory submission types and processes the systems support using the data. Commonalities among different data domains of CBER's regulatory administrative databases are discussed. These commonalities have evolved enough to constitute real database convergence and provide a valuable asset for business process intelligence. Balancing review workload across staff, exploring areas of risk in review capacity, process improvement, and presenting a clear and comprehensive landscape of review obligations are just some of the opportunities of such intelligence. This convergence has been occurring in the presence of usual forces that tend to drive information technology (IT) systems development toward separate stovepipes and data silos. CBER has achieved a significant level of convergence through a gradual process, using a clear goal, agreed upon development practices, and transparency of database objects, rather than through a single, discrete project or IT vendor solution. This approach offers a path forward for FDA systems toward a unified database.

KEY WORDS: database; FDA; managed review; PDUFA; regulatory workload; silos.

BACKGROUND

The Center for Biologics Evaluation and Research (CBER) is the Food and Drug Administration (FDA) center with regulatory oversight of biological products including vaccines, blood and blood products, and cells, tissues, and gene therapies for the prevention, diagnosis, and treatment of human diseases, conditions, or injury (2). This makes it unique among FDA centers and challenging for the variety of areas of responsibility and authority under which it must operate (3). The diverse array of products are subject to an equally diverse array of regulations, statutes, and negotiated user fee performance goals. In addition, these regulations and performance goals vary across the product lifecycle, from investigational through post-marketing and surveillance. It must therefore receive and process a variety of regulatory applications. For these reasons, the databases that support

these applications and their associated review processes can be quite complex.

These systems not only must track the actual submissions CBER receives, but their metadata such as what is being added or modified in the product or manufacture, workflow, review committee membership and roles, review decisions, precedence, internal and external communications, and details surrounding the product, manufacture, usages, dosage forms, interfaces to the documents, and submissions in CBER's electronic document room, as well as history.

CBER's authority, actions it can take, reporting obligations, and how it must manage the review process referred to as "Managed Review" differs across product classes and application type as do the content and structure of the applications themselves. These include the Biologics License Application (BLA) under Title 21 of the Code of Federal Regulations parts 600 and 601 (21CFR600s), New Drug Application (NDA) and Abbreviated New Drug Application (NDA) under 21CFR314, Pre-Market Notification under section 510(k) of PHS Act and 21CFR807.81, Pre-Market Application (PMA) under 21CFR814, Investigational New Drug Application (IND) under 21CFR312, Investigational Device Exemption (IDE)

¹ U.S. Food and Drug Administration (FDA), Rockville, Maryland 20852, USA.

² To whom correspondence should be addressed. (e-mail: JeffreyK.Smith@fda.hhs.gov)

under 21CFR812, Lot Release under 21CFR610, Emergency Use Authorization (EUA) under section 564 of the FD&C Act, and formal meetings with industry under section 119(a) of the FDA Modernization Act (FDAMA). In addition, user fee performance agreements established with regulated industry have created a disparate set of review timelines and reporting obligations across application types, *e.g.*, the Prescription Drug User Fee Act (PDUFA), Medical Device User Fee and Modernization Act (MDUFA), and the Food and Drug Administration Amendments Act (FDAAA).

This diverse environment presents a challenge not only in the management and tracking of regulatory applications and review but also in the development of IT systems that support them. The focus here is the databases underlying the supporting systems that capture metadata about the regulatory applications and support the review process.

CHARACTERIZATION OF SYSTEMS AND THEIR UNDERLYING DATABASES

There are seven major IT systems that support regulatory applications tracking and review and all pertain to this discussion.

- Regulatory Management System for the Biologics License Application (RMS/BLA) for the BLA marketing applications
- Blood Logging and Tracking for the PMA, NDA, ANDA, and 510(k).
- Biologics Investigational and Research Application Management System (BIRAMS) for the IND, IDE, Master Files, and EUA.
- CBER Regulatory Meetings Tracking System for information about formal meetings with regulated industry.
- Document Accountability and Tracking System (DATS and DTS) for internal document tracking and routing of both paper and electronic application documents.
- Lot Release System for support of information about product lots and release protocols in support of approval and for routine surveillance.
- Pre-Application Tracking System for support of information related to submissions submitted by sponsors to obtain feedback from CBER prior to

submitting a formal investigational or marketing application, *e.g.*, *pre-IND*, *pre-BLA*.

Figure 1 presents these regulatory systems in the context of the pharmaceutical product lifecycle they support, from investigational through post-market.

The databases can be quite complex. RMS/BLA alone, for example, fulfills more than a thousand documented user requirements and is comprised of more than 35 major screens, 40 queries, and 50 reports. The physical database consists of about 200 tables, 350 triggers, 40 packages, and 20 views.

The data underlying these systems reside within the same Oracle database instance, but are partitioned into separate schema accounts. An Oracle schema is a grouping of tables and other database objects by the IT system or account that owns them. The schemas were designed to be transactional, for efficient data insertion and storage as opposed to reporting. Therefore, the data models are relational and normalized, meaning that the data are spread across multiple, smaller tables, to minimize redundancy, that can then be joined together (related) in a variety of ways using common keys. However, the degree of normalization varies across schemas. Each system has reporting capability within its user interface, *i.e.*, screens and reports, against its respective data domains and is specific to only those regulatory applications it supports.

CBER uses some popular tools for business intelligence (BI) reporting, such as Business Objects. These tools are built around a datamart principle whereby the data from the live, transactional systems are decoded, aggregated, and pieced together into simpler structures (transformed) and then ported into a separate database for reporting. The reporting database must then be refreshed at regular intervals. This principle facilitates data mining and analysis and is often referred to as “On-Line Analytical Processing” (OLAP). At present, the usage of OLAP in CBER is limited mostly to adverse events reporting and analysis. What does exist that is related to the regulatory application systems is confined to each system and regulatory application type. This tool and approach provide much value to CBER but, by its very nature, would not demonstrate convergence of the transactional databases themselves that underlie the different IT systems. Further, government agencies are beginning to look beyond basic data warehousing and reporting toward Master Data Management (MDM) strategy which will be discussed (4).

Product Development Lifecycle

Fig. 1. CBER regulatory systems pertaining to this article in the context of the phases of the pharmaceutical product lifecycle they support

SILOS BY DESIGN

It is not unusual for organizations to find themselves with multiple and disparate information systems. Such an outcome often breeds stovepipes, or more accurately, isolated silos of data (5). This predicament has been attributed to a variety of reasons, most often a command-and-control-oriented culture (Rosen) and IT governance (6).

However, organizations arrive at such a state for practical reasons as well. As CBER has evolved over the years, so too have the regulations, statutes, agreements, and the IT systems that support its mission. Separate IT

systems supporting different areas of need were a natural outcome as regulatory application types, and the organization itself were, at one time, partitioned along similar lines of separation. Separate databases underlying these systems were developed and modified over time, as needs arose, by different sets of stakeholders working in separate organizational units, drawing from different sets of requirements, all working to meet different deadlines. IT system development in CBER conforms most closely to the “spiral model,” a continuous cycle of enhancement and deployment that has proceeded along different timelines for the aforementioned reasons but also because of the nature of funding sources.

Targeted funding brought needed resources but also brought challenges to an otherwise, deliberative, unified design approach. Funding is often highly targeted as the legislation behind each funding source specified areas to be addressed and tracking and reporting requirements, and drove development schedules. This all but precluded the use of a fund beyond its immediate mandate. Even with strong IT governance, development in one domain could not await the development in another when one had been legislated and already funded. Therefore, separate silos of data are often an unavoidable outcome of such a dynamic environment.

Some important legislations that have driven FDA IT systems development in this way are the PDUFA of 1992 and its subsequent amendments, FDAMA of 1997, Bioterrorism Act of 2001, Project Bioshield of 2004, MDUFMA of 2004, Pandemic and All-Hazards Preparedness Act of 2006, Best Pharmaceuticals for Children Act of 2007, Pediatric Research Equity Act of 2007—the latter two being enacted under FDAAA.

To fulfill PDUFA, for example, it would have been unacceptable to use the accompanying funds on IT systems or data domains not covered under the Act, such as whole blood and blood components for transfusion, devices, and compliance activities. Project Bioshield specified funding for implementation of the Emergency Use Authorization (EUA). It would have been unacceptable for delays as part of a larger unified database vision to delay support for the EUA.

As a federal agency, FDA is also impacted by other external legislation such as the Healthcare Insurance Portability Act of 1997, Federal Information Security Management Act of 2002, American Recovery and Reinvestment Act of 2009, as well as Federal Records Acts.

CONVERGENCE

Reorganizations gradually brought together different stakeholders as they mixed together different areas of regulatory application review. Staff found themselves managing multiple application types and having to navigate across multiple IT systems. Differences among these systems became easily apparent and troublesome. Realization of this prompted the CBER vision of a Vast Regulatory Database in 1997. Several outputs of this vision have endured and have gradually driven the underlying databases toward a significant degree of

harmonization, even as FDA and IT staff have turned over and development has continued independently within each IT system team. Development under the spiral model enabled these outputs to be gradually introduced with successive versions of each system. Although greater harmonization is still desired, enough has occurred in key areas to provide important value to CBER for review management and business intelligence.

Notably, some of the key areas are consistent with entities described in Oracle's MDM (7), which may help explain their value to CBER. Commonalities now exist across a number of data domains, database objects, and architecture to substantiate convergence and therefore real progress toward the vision of a unified database.

Database convergence in CBER is evident in a number of areas. These are:

- The data underlying all CBER regulatory IT systems reside within the same database instance, although partitioned into separate schema accounts.
- Broad use of shared database objects by multiple schemas, *e.g.*, tables, entities or columns, keys, values, stored procedures and functions, sequences
- Shared use by all systems of a common schema account
- Sharing/harmony of some key data definitions
- Commonalities in the storage and characterization of critical dates across schemas
- Similarities in logical architecture
- Similarities in scheme administration
- Use of public synonyms

This was accomplished largely because development teams agreed to make their objects visible (transparent) to each other, exercise a common naming convention for their objects, format similar data types similarly, and publicize useful code for reuse. There are still profound differences among database schemas, *e.g.*, varying degrees of normalization (some highly relational and codified) and differing data domain models. However, the aforementioned master data nature of the commonalities minimizes the importance of these differences with respect to business intelligence as the MDM model suggests. For example, shared tables exist for valid values and pick lists, reviewer names, review organizations, review schedule types, company contacts, manufacturing facilities and addresses, and product names. Critical dates are stored and named in similar ways such as dates for receipt, review status, and completions. Due dates are stored physically different across schemas but logically similar, *e.g.*, either as values in physical tables (in RMS/BLA) or as derived values encapsulated in database objects such as a view (in BIRAMS). These objects are similarly named across schemas and kept conspicuous with public synonyms. A submission numbering and hierarchy scheme ensures unique numbering and parent-child relationships across application databases, the submission tracking number (STN).

All database schemas underlying these various IT systems reside within the same database instance. This allowed for greater transparency, that is, an awareness and

understanding of useful database objects by the developers of the different systems that could then be shared among them. Reuse of existing objects, such as tables, views, and functions, is often pursued as the path of least resistance when developers are aware such assets exist (8). This “object transparency” facilitated the shared use in CBER of existing database objects by the different development teams.

Establishing master data is not without its challenges. For example, some data domains that would provide great value as master data remain a challenge. Legal entity continues to be difficult to harmonize across databases. Entities such as sponsors of INDs can be a person or an institution and can be different than an applicant of a new drug application, or a contract facility, or the manufacturer of its source material. These legal entities often merge, split, or are renamed, and some are subsidiaries to others. Their applications are often submitted to FDA using variations of the same entity name. Therefore, harmonizing and managing these across departments, application types, and systems is still a work in progress.

AFFIRMATION

Database convergence is becoming evident from a growing number of uses that are benefiting CBER in new ways. Queries written directly against the live, transactional data, across schemas, use a number of the above master data commonalities. These offer greater BI than the traditional reports within each specialized IT system.

The following sample Structured Query Language (SQL) code is used within the database and is evidence of MDM. It fetches the complete review committee list for each IND application from the BIRAMS database schema using master data. The master data here are actually a function which selects against a shared table of Center review staff, the CBER *person* table. Such functions are provided within each database application schema so these essential data can be derived in the same way across the different regulatory systems. Figure 2 presents the query as an execution diagram.

```
SELECT DISTINCT folder_id stn, f_get_ind_reviewers(folder_id) committee
FROM [BIRAMS schema].ira_data_vw;
```

The function *f_get_ind_reviewers* is used by the query and makes use of the master data, *person*. Note its use in an excerpt of the function's code below.

```
SELECT DISTINCT RTRIM(per.name_last||', '||per.name_first||' '||per.name_mi) reviewer_name
FROM ira_reviewers ind,
[shared schema].person per
WHERE ind.folder_id = p_ind_id AND ...
```

**Actual schema names, in brackets, are not displayed here for security reasons.*

A comprehensive review workload can be derived across schemas (across databases) by making use of some of the master data objects discussed above (Fig. 3) (data presented here are fictional). For each reviewer, the graph in Fig. 3 presents a composite bar of the number of submissions currently under their review by application type. In the example, it appears that reviewers L and N have the greatest workload followed by reviewers K and M. Another variation of this workload tool is also employed which adds a time dimension, *i.e.*, workload by reviewer by month that their submissions are coming due for a regulatory action (not shown here).

However, some application types will require more effort to review than others, *e.g.*, an original BLA or NDA *versus* an annual report. Weightings can be applied by application type to further aid in gauging workload. Just what each numeric weighting should be is not the subject of this paper but would be subject to the knowledge and discretion of a frontline supervisor and will vary across review disciplines, *e.g.*, clinical

review, promotional materials review, regulatory review. Weightings were applied to the sample data from above giving some application types a greater factor than others (Fig. 4). Reviewer M now appears to have the greatest workload followed by reviewers N than L. More importantly, however, this reveals that 70% of the overall weighted review workload is being handled by 32% of the review staff. Such a disproportionate dependency upon so few reviewers suggests a risk in review performance. It can inform decisions of work redistribution, cross-training, and help in gauging management effectiveness in the distribution of work across their staff. Such a model can also be used for comparisons with future distributions for planning and management performance. This tool is currently being employed within the Clinical Review Branch and Regulatory Project Management Branch of CBER's Office of Blood Research and Review.

Applying another dimension, product, provides another way to detect risk to the review process (Fig. 5). Figure 5 presents the number of applications received by product within a specified period of time (cohort) against the number of unique product reviewers serving on the committees for

Fig. 2. Execution diagram of SQL query using master data objects—a function and a table of persons

those submissions. Such a presentation can help inform in the adjustment of review capacity to where it is most needed, to those product areas with most applicant submission activity. It can also expose risk in review capacity or understaffing in a product area. In the example, products x, y, and z may be good candidates for more review staff participation since only a few reviewers were associated with these products that receive a disproportionately higher number of submissions (STNs are a measure of submission counts).

Another output that makes use of master data and that has been useful to CBER is a comprehensive list of review obligations and due dates across application databases, example not shown here. Such reports are up to the minute because they run against the live, transactional databases. Such queries are used regularly to ensure managers have complete visibility of all pending applications and their due dates but have also been used on multiple occasions in preparation for snow storms and government shutdowns.

Fig. 3. Bar graph depicting the count of each application type from disparate databases then distributed across review staff. Such a presentation allows a manager to distribute work in an informed way by more than mere counts but by type of applications. Application types which vary in their effort and time to review

Fig. 4. Reviewer workload is derived across disparate databases as a count of submissions weighted by submission type and then arrayed as a normal distribution. Seventy percent of overall weighted workload is determined to be held by 32% of review staff

DISCUSSION AND CONCLUSIONS

For large, disparate database systems in FDA, unification through gradual convergence can be accomplished to a degree that is sufficient enough to facilitate real-time business intelligence. This may be best approached as a common goal, using agreed upon standards and practices and by focusing on master data. These practices and standards include identifying data domains as master data targets, object transparency, common naming conventions, code publication and reuse. This approach

allows systems to continue to evolve independently and flexibly yet still converge using the data domains that count, master data. Such an approach may become especially important as organizations migrate toward new cloud services. Former Health and Human Services' Chief Information Officer John Teeter noted that “The danger of the cloud is that it could just move our stovepipes to a more accessible environment” (9). Gradual convergence using master data may be a path forward whereby each IT system can remain responsive to its unique and changing mandates and continue to support FDA's mission.

Fig. 5. Count of regulatory submissions received, by product, against the number of distinct reviewers participating in the review of those submissions. Such a model can expose a risk in review capacity or an underutilization of reviewers for products where FDA is receiving a growing number of submissions

Notice The findings and conclusions in this article have not been formally disseminated by the Food and Drug Administration and should not be construed to represent any Agency determination or policy. All data presented in this paper are for demonstration purpose and are fictional. Actual schema names have been redacted and substituted in brackets.

REFERENCES

1. Buckley CA. Leveraging regulatory information and global process standards. DIA 2012. Presentation at the DIA Conference on EDM and ERS/eCTD; 2012 Oct 9–11; Baltimore. <http://www.diahome.org/en/Flagship-Meetings/12003-EDM.aspx>. Accessed 10 Nov 2012.
2. Food and Drug Administration [Internet]. Rockville (MD): CBER Vision and Mission; c2005 [updated 2009 Jun 18; cited 2012 Nov 10] <http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CBER/ucm122878.htm>. Accessed 10 Nov 2012.
3. Sensabaugh SM. A primer on CBER's regulatory review structure and processes. *Drug Inf J*. 1998;32:1011–30. http://www.diahome.org/productfiles/8357/diaj_12632.pdf. Accessed 12 Aug 2012.
4. Romeo J. Getting data silos under control. *Government Computer News*. 2007 Aug 3. http://gcn.com/articles/2007/08/03/getting-data-silos-under-control.aspx?sc_lang=en. Accessed 12 Aug 2012.
5. Rosen E. Smashing silos—five steps to encourage collaboration and do away with insular business units acting at cross-purposes. *Bloomberg Businessweek*. 2010 Feb 5. http://www.businessweek.com/managing/content/feb2010/ca2010025_358633.htm. Accessed 12 Aug 2012.
6. Gillispie J, *et al*. Data governance—managing information as an enterprise asset. Lexington (KY): National Association of State Chief Information Officers (NASCIO); 2008 Apr. <http://www.nascio.org/publications/documents/NASCIO-DataGovernance-Part1.pdf>. Accessed 12 Aug 2012.
7. Butler D. Better information through master data management—MDM as a foundation for BI. Oracle White Paper. Jul 2010. <http://www.oracle.com/us/products/applications/master-data-management/018874.pdf>. Accessed 12 Aug 2012.
8. Özyer T, Kianmehr K, Tan M. Recent trends in information reuse and integration. New York: Springer Science + Business Media; 2012. ISBN 978-3-7091-0737-9.
9. Walker MB. CIOs risk recreating data silos in the cloud. *Fierce Government IT*. 2011 Aug 24. <http://www.fiercегovernmentit.com/story/cios-risk-recreating-data-silos-cloud/2011-08-24>. Accessed 12 Aug 2012.