Bioinformatics Modernization and the Critical Path to Improved Benefit-Risk Assessment of Drugs

Armando Oliva, MD
Deputy Director for Bioinformatics, Office of Critical Path Programs

Randy Levin, MD
Director for Bioinformatics, Office of Critical Path Programs

Rachel Behrman, MD
Director, Office of Critical Path Programs

Janet Woodcock, MD
Chief Medical Officer and Deputy Commissioner, Office of the Commissioner, Food and Drug Administration, Rockville, MD

At FDA, bioinformatics *means the design, development, and use of modern computer systems to efficiently and effectively manage the* regulatory product information supply chain, *along which medical product information travels among many relevant organizations. The FDA relies on efficient management of this information to assess a drug's safety and effectiveness. The current bioinformatics infrastructure that supports product information exchange is inefficient and is comparable to the antiquated infrastructure of the financial industry in years past. Bioinformatics modernization requires improvements in three important information management domains: access, standards, and interface. We must have better access to information, more standardized information, and better interface with information (ie, better tools to convert information into knowledge). The FDA has taken measurable steps to modernize its bioinformatics infrastructure, but the effort is costly, complex, and time consuming. Nonetheless, it is a necessary step to improve the Critical Path and enhance benefit-risk assessments of drugs.*

Key Words
FDA; Bioinformatics; Information management; Benefit-risk assessment; Critical Path

Correspondence Address
Armando Oliva, MD, 5600 Fishers Lane HF-18, Rockville, MD 20857 (email: armando.oliva@fda.hhs.gov).

Presented at the 25th annual DIA Clinical Data Management Meeting, Medical Informatics Opportunities to Improve the Benefit-Risk Assessment of Drugs, March 19, 2007, 1:30–3 PM. The views presented are those of the authors and not necessarily those of the Food and Drug Administration.

INTRODUCTION

Bioinformatics is the design and development of computer-based technology to support the life sciences (1). At FDA, bioinformatics means the design, development, and use of modern computer systems to efficiently and effectively manage the *regulatory product information supply chain*, along which medical product information travels among the many relevant organizations (eg, study data travel from applicants to FDA for product review; adverse events reports travel from industry and the public to FDA; drug prescribing information travels from FDA to the public). FDA is a critical link in this information supply chain, and ensuring the efficient exchange of drug information between FDA and its stakeholders is critical to FDA's mission to protect and promote public health and to improve risk-benefit assessments of drugs (see Figure 1).

FDA has made real strides in its effort to achieve a fully automated and interoperable infrastructure for managing the exchange of regulatory product information. For example:

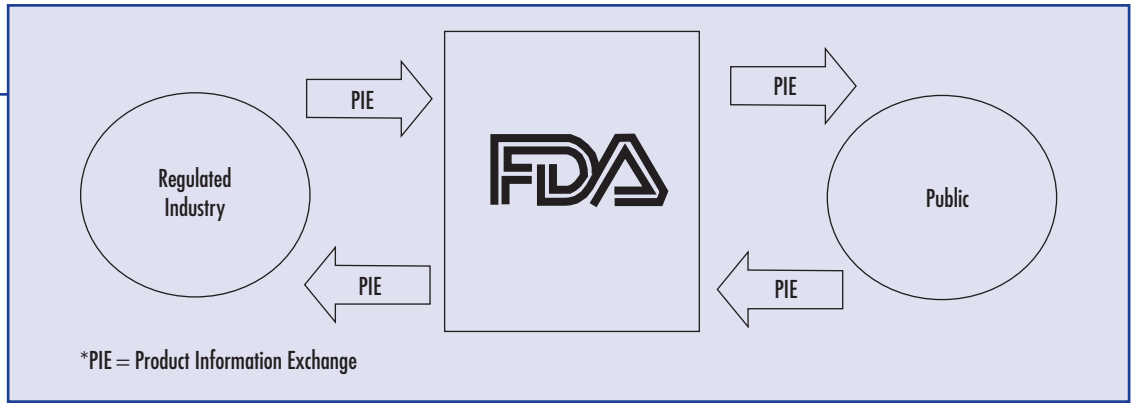
- FDA now has a single portal through which product information reaches the agency electronically.
- Prescription drug labeling information must be submitted to FDA electronically in a standardized format.
- Several regulations have been proposed, or will be proposed soon, to require the electronic listing of products and the electronic registration of manufacturing establishments as well as the electronic submission of other regulatory information.

However, inefficiencies in the regulatory product information supply chain remain, and additional resources are needed to continue this long-term effort to modernize the agency's bioinformatics infrastructure for the 21st century.

Through its Critical Path Initiative, launched in 2004, FDA is seeking to stimulate and facilitate a national effort to modernize the scientific process through which a potential human drug, biologic product, or medical device is transformed from a discovery or proof of concept into a medical product (2). In March 2006, FDA issued the *Critical Path Opportunities Report* (3). This report identifies specific areas in product development where improvements can increase efficiency, predictability, and productivity in the development of new medical products. One of these areas is bioinformatics. Modernizing FDA's bioinformatics infrastructure has many critical benefits. For example, a modern bioinformatics environment will:

Submitted for Publication: May 30, 2007
Accepted for Publication: October 10, 2007

FIGURE 1



The regulatory product information supply chain. The regulated industry gives FDA certain information about a product (eg, new drug application). The FDA evaluates that information and, if the product is approved, makes certain product information available to the public (eg, drug labeling). The public in turn provides product information to the agency (eg, adverse event reports), which FDA processes, providing updates to the regulated industry (eg, recall) and the public if needed (eg, new labeling, warning box).

- Enhance FDA regulatory decision making
- Improve the safe use of human and veterinary drugs
- Get safe and effective medical devices to patients faster
- Improve FDA's ability to evaluate the safety and nutrition of food and food ingredients
- Improve FDA's ability to detect and mitigate contaminations in the nation's food supply

Effective bioinformatics modernization requires understanding the existing problems in today's infrastructure and addressing them. This article describes FDA's vision to harness bioinformatics within the agency to support its public health mission. Progress in this area will also ultimately help modernize the Critical Path sciences.

THE BIOINFORMATICS PROBLEM

To understand the current inefficiencies in the regulatory product information supply chain and to fully appreciate the extent to which change is needed, one must only think back to the state of the financial industry in the earliest days of the nation, before the creation of today's modern financial information infrastructure. Back then, we lacked a standard currency, and wealth was maintained in a variety of different forms (eg, gold, silver, beads, animal pelts). There were few banks, if any, and no other places to store one's wealth, so individual wealth was commonly stored under a mattress or hidden under floorboards. Money was counted by hand and transported in sacks or bags to its destination. Looking back from today's perspective, it is easy to imagine the inherent inefficiencies of such a system. How difficult (or impossible) were the simplest financial transactions that we now take for granted? Today, the financial industry has a standard currency (US dollars), has improved access to money and transaction methods (eg, banks, ATMs, credit and debit cards, financial networks), and has developed user-friendly interfaces with financial information (eg, automated money-counting machines, calculators, the Internet). Our modern financial infrastructure allows money to travel quickly and securely throughout the country (and the world), and financial information is accessible virtually anywhere. This modern financial infrastructure has enabled the development of new interfaces with our financial information that were previously unthinkable, making the information increasingly useful both to the nation and the individual. For example, the Federal Reserve now has tools to monitor billions of financial transactions almost in real time to assess the health of the economy and make better monetary policy decisions. For the individual user, certain banks now provide instant messages to an account holder's cell phone when the account balance dips below a certain amount.

Our health information infrastructure today is on a par with the finance industry of years past. Using the analogy that data collected during a clinical study for a new drug is money, we have no widely implemented standard currency for study data. Furthermore, companies store their

study data locally and in different formats, making it extremely difficult to exchange data quickly and securely from one entity to another. Available tools to analyze study data are difficult to use, making the assessment of a drug's safety and efficacy difficult and time consuming.

The submission of premarketing applications is a good example of the existing inefficiencies in the regulatory product information supply chain. Although many companies are now using automated technologies to collect, sort, and organize their new drug information and some applicants send FDA electronic submissions, some still send paper. Among those who send product information electronically, it may come in scanned portable document format (PDF) that is not computer readable.

Although electronic submissions continue to improve, many of us still use different technologies, different computer languages, and different terminologies, making communication of information inefficient and sometimes rendering the information incomprehensible. If that were not problematic enough, within FDA, systems, technologies, and terminologies often differ or are duplicative. We have multiple submission formats, and much of the information we receive is still entered manually into various computer systems. For example, some applications can be all paper, mixed paper and electronic, or all electronic. Often it is not clear what information is provided in what media type, and this differs from submission to submission. The resulting inefficiencies and systems redundancies mean increased costs for human and technological resources, making it difficult for FDA to carry out its mission effectively. Unlike the Federal Reserve's ability to leverage our modern financial information infrastructure to inform monetary policy, the lack of a modern bioinformatics infrastructure hinders FDA's ability to make better health policy decisions about new drugs.

We must maximize the usefulness of the information we collect and share—the multitude of data gathered during research and during clinical studies that are packaged and sent to the FDA for marketing approval and the data we gather during postmarketing surveillance. To do this, the FDA and its regulated industry must undergo a major transformation. By implementing a standard currency for the data, improving access to the data, developing user-friendly tools (ie, interfaces) that convert data into knowledge, and ensuring that these changes occur throughout the regulatory product information supply chain, we can lower development costs, shorten the time to market for new drugs, and enhance our ability to communicate information about the safe and effective use of FDA-regulated products, thus promoting the public health. FDA recognizes its leadership role in this transformation.

THE SOLUTION

For years, FDA has been working toward developing a modern bioinformatics environment. Achieving this goal will enable the agency to more efficiently and effectively use the information it receives from regulated industry and the public to improve benefit-risk assessments, enhance regulatory decision making, and communicate its findings to all stakeholders. In addition, our activities will facilitate ongoing efforts by FDA stakeholders to modernize their health information management systems, to enhance product information exchange among each other and with FDA, and to improve their decision making. In the clinic, this means improving treatment decisions at the bedside.

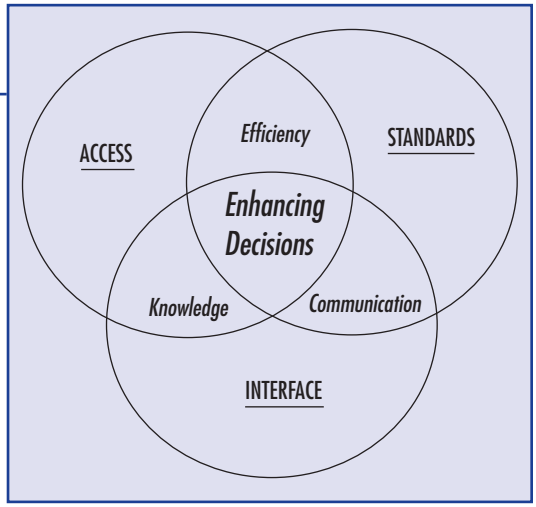
A modern bioinformatics environment requires enhancing and integrating three key information management domains: (1) information standards, (2) information access, and (3) interface tools that can efficiently convert information into knowledge. Standards, access, and user-friendly interface tools work together to influence the way we receive, manage, and communicate information as part of regulatory decision making (see Figure 2).

STANDARDS

By developing, adopting, and implementing common standards, we are establishing a common language for managing and exchanging health-related information, including research

FIGURE 2

Three information management domains.



and study data. Standards enable both people and electronic information management systems to communicate effectively and efficiently. FDA's regulations have established some standards. For example, all communication with the agency must be in English. However, we must use a standard language that both humans and computer systems can read to ensure the smooth exchange of information along the regulatory product information supply chain to automate the processing of information. We must also agree on standard terminology: formal and consistent definitions as well as examples so that we can effectively share concepts and ideas. Computer systems have their own language with precise terminologies and dictionaries (ie, semantics) and grammar (ie, syntax). Using precise, standard language, including standard syntax and semantics, all along the regulatory product information supply chain is a must for creating an effective bioinformatics environment.

ACCESS

Improving access means developing tools that make it easy to send, receive, and share information. As with standards, our regulations address some information access issues. For example, certain information must be sent directly to FDA so we can access it whenever needed (eg, marketing applications). Other information (eg, medical records collected during a trial) must be maintained at specific off-site locations and made available to the FDA at certain times. Mechanisms to exchange information electronically also improve access. When the systems are functioning properly, electronic mail is accessible 24 hours a day, seven days a week, at just about any location. This accessibility is far superior to paper mail. The same information sent via electronic mail often arrives at its destination in seconds or minutes instead of days.

Accessibility goes beyond having access to electronic files or paper documents, however. For FDA, access is also measured in terms of our ability to quickly get to specific information contained in files and documents. It is not sufficient to easily access a review of a new drug application; we also need quick access to detailed information within a review, such as information about an unusual adverse event. We must be able to "computerize institutional memory," to better manage knowledge as well as information.

INTERFACE

The third key information management domain is interface. The usefulness of information is greatly diminished if the appropriate interfaces are unavailable. We must design and make widely available user-friendly tools for analyzing information that efficiently and effectively convert information into knowledge. We must also make sure our users are properly trained to use these tools. FDA regulations address some of these interface needs. For example, today, a marketing application must contain a comprehensive table of contents, a tool for finding information quickly. Increasingly, the modern interface tools we use are software programs on a computer.

These three domains—standards, access and interface—work interdependently to improve decision making. Even if information is more accessible, without a standard language, our efficiency will be hampered. If we use common standards, but our interface tools are inadequate or difficult to use, we will have difficulty communicating. Finally, common standards and user-friendly interface tools cannot overcome a lack of accessibility. Improvements in all three key information management domains will be

needed to enhance decision making and improve the public health. Conversely, problems in any one area can significantly impede decision making.

Creation of a national bioinformatics infrastructure is not something industry and other relevant stakeholders can achieve without federal leadership—the government has already taken the lead by launching its broad Health Information Technology (IT) effort (4). Similarly, modernization of the regulatory product information supply chain cannot be effected by industry alone. In its Critical Path Initiative, FDA identified harnessing bioinformatics as one of the five areas needing most work (5). By harnessing bioinformatics to modernize our own information management environment, FDA will help pave the way for its stakeholders to do the same.

It is FDA's goal to achieve a modern bioinformatics environment that enhances its decision-making capability and facilitates product information exchange. Ideally, this means that the information flowing through the regulatory product information supply chain is standardized, both information and knowledge are accessible electronically, and the tools for processing, analyzing, and reviewing information are both effective and user-friendly.

Figure 3 is a conceptual model of a possible modern bioinformatics environment for FDA and its stakeholders. Regulatory product information is accessible in an all-electronic integrated system consisting of repositories for electronic documents and data. The information in the repository is managed using detailed, standardized information about each file in the repository. Information based on established standards is entered into the system (IN) or taken out of the system (OUT) through user-friendly tools. (Such a system ultimately could even be managed by a third party, reducing the burden to FDA and stakeholders by reducing the need for redundant systems.)

NEXT STEPS

FDA has started down the modernization path by taking calculated steps to achieve measurable and incremental improvements in its decision-making capability and in product information exchange. We are working to ensure that each step also contributes to one or more of the three key information management domains described previously. Although each step is a stand-alone effort, FDA has been engaging stakeholders throughout to ensure that each step aligns with related efforts within regulated industry and with other federal health IT efforts. We also have tried to maximize available resources by minimizing duplication of effort. Whenever possible, resources have been leveraged by, for example, using or upgrading existing systems; forming partnerships within FDA centers and outside the FDA, including other government agencies (eg, the National Cancer Institute), nongovernment organizations, and

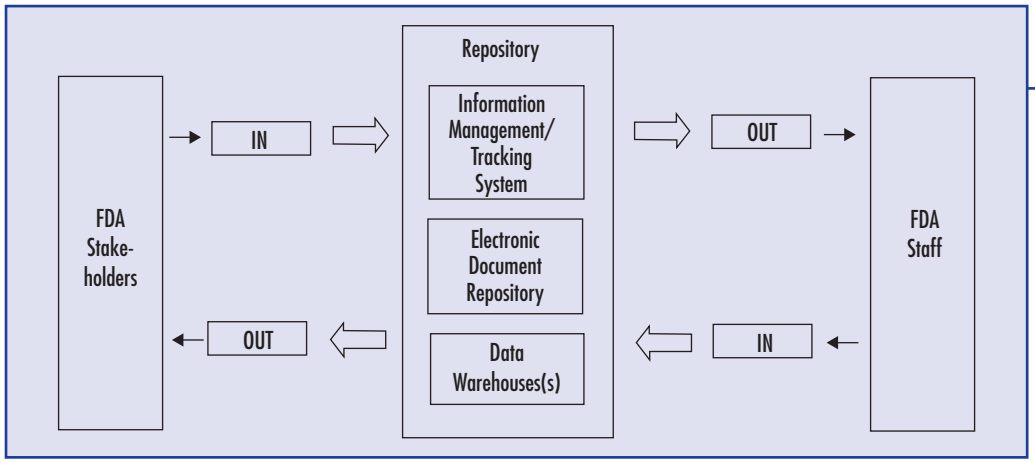


FIGURE 3

The modern bioinformatics environment.

vendors; and adapting systems developed to manage one type of document to be able to manage other document types.

During the next five years, FDA hopes to lay a solid foundation for its modern bioinformatics environment. We have identified specific key IT initiatives for completion, including the following:

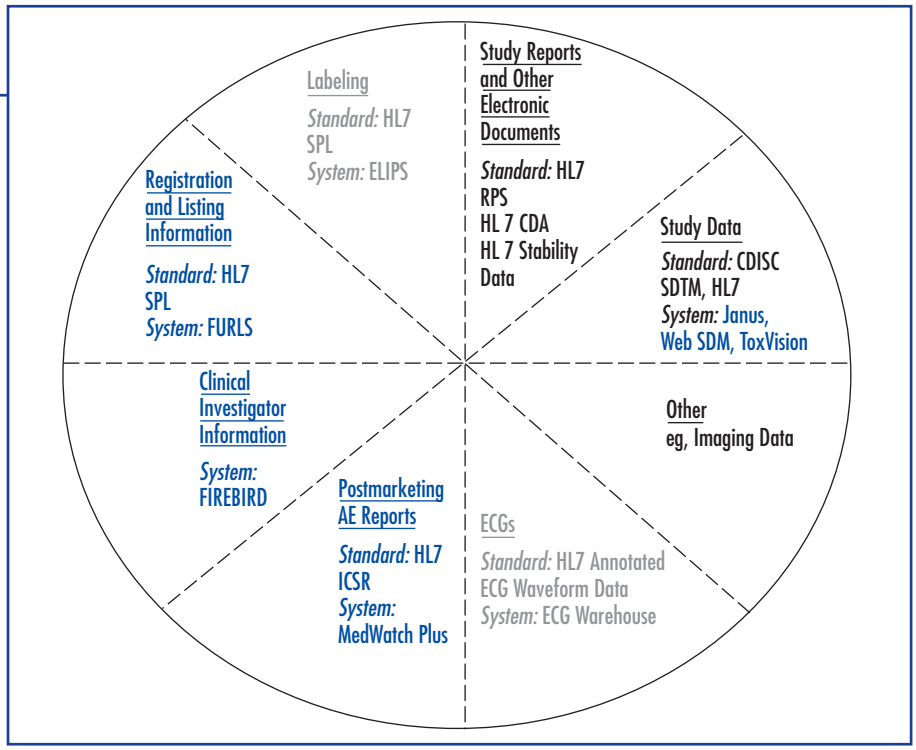
- Expanding agency-level coordination and governance of FDA's overall bioinformatics effort (ie, the FDA Bioinformatics Board was created in early 2006; see discussion below).
- Introducing solid information management concepts into ongoing specific standards, access, and interface development activities: for example, development and adoption of HL7 exchange standards such as structured product labeling (SPL) and the regulated product submission (RPS) standard; enhancement of the electronic submission gateway (ESG) to facilitate bidirectional exchange of product information; various cooperative research and development agreements (CRADAs) to develop user-friendly review tools.

These initiatives promise a high return on investment. Figure 4 graphically depicts the specific projects that have been identified for completion. Some have been launched, some are still under development, and some are in the planning stage. Each piece of PIE (product information exchange) is handled as an individual project, with its attendant regulatory and policy requirements. For each piece, FDA has worked, or is working, with relevant stakeholders and accredited standards development organizations to develop and adopt specific standards and develop and implement the piece.

A significant constraint in achieving a modern bioinformatics environment is the lack of adequate resources, both human and fiscal. We need more personnel with expertise in business process automation to develop requirements and terminology. We need project managers and technology experts who can help develop and implement the computer systems that will meet those business requirements. We also need better training so we can use modern computer systems to their maximum benefit. Adequate and consistent fiscal support also is needed to acquire the requisite human resources as well as equipment acquisition (eg, hardware and software). In the past, public-private partnerships have been particularly helpful in lowering FDA

FIGURE 4

FDA's product information exchange (PIE) projects. Projects in gray have been implemented; projects in blue are at some stage of development or implementation; projects in black are in planning.



development and implementation costs, providing basic services to FDA and its stakeholders, and offering certain additional fee-based value-added services.

Another ongoing constraint is that FDA continues to receive a significant amount of information on paper. As described previously, paper greatly limits our access to information. FDA has been encouraging submission of electronic information for many years, but we need an all-electronic information environment. A modern bioinformatics infrastructure cannot realistically be achieved as long as information is stored on paper.

To oversee bioinformatics modernization efforts, FDA established its Bioinformatics Board on February 21, 2006. Its goal is to provide agency-level coordination and governance of all bioinformatics efforts leading to the creation of a modern bioinformatics environment. The Bioinformatics Board's mission is to oversee the planning and management of FDA's information management and bioinformatics activities to move in a coordinated manner toward a highly automated, mission-supportive, information management environment. The board will increase the pace of technology standardization to the extent feasible and strive for the most efficient and effective use of resources across the agency. The bioinformatics modernization effort is costly, complex, and time consuming. Nonetheless, it is a necessary step to improve the Critical Path and enhance benefit-risk assessments of drugs.

REFERENCES

1. Lacroix Z, Critchlow T. *Bioinformatics: Managing Scientific Data*. San Francisco: Morgan Kaufman; 2003.
2. See http://www.fda.gov/oc/initiatives/criticalpath/.
3. *Critical Path Opportunities List and Report*, March 2006. Available at: http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_report.pdf.
4. See http://www.hhs.gov/healthit/.
5. *Critical Path Opportunities List and Report*, March 2006, Topic 3: Harnessing Bioinformatics, p. R-15. Available at: http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_report.pdf.

The authors report no relationships to disclose.