The Next Phase of Industry Evolution: Driving Value from Big Data in the Life Sciences

CUSTOM INDUSTRY BRIEF
Sponsored by: Hitachi Data Systems

Eric Newmark
June 2012

IDC HEALTH INSIGHTS OPINION

Life science companies face data-related challenges across all facets of the business. The industry has long been afflicted by its affinity for large data silos distributed across disparate environments, but those data sets are growing at rates that will soon surpass the capabilities of present infrastructure, rendering some current approaches to data management obsolete. At the same time, the life sciences are diligently working to break down information barriers (both internal and external) to promote better data sharing and collaboration across divisions and partners. As silos are eliminated and massive data sets are merged, these newly combined repositories of both structured and unstructured data will require much greater storage, search, and analytical power to support ongoing business initiatives and future innovations. This reality exists not only in R&D but across all areas of the enterprise. The current data explosion is true for sales, marketing, manufacturing, supply chain, research and development (R&D), and even various outsourcing and collaborative partnerships.

Clearly big data is not just about volume, as has been painstakingly documented by the four Vs (volume, velocity, variety, and value). As such, it is vital that infrastructure remain strong enough to sustain not just big data but high-performing content. Data platforms must provide scalability that will enable life science companies to capture, search, and perform real-time analysis on petabytes of information throughout the life cycle of the data. Implementing appropriate infrastructure to ensure this happens is among the most important investments that life science companies need to make.

Industry Background

The life sciences have undergone significant transition for most of the past decade. As the pharmaceutical industry’s patent cliff has grown increasingly dire, with more than 20% of industry revenue now exposed or at risk in the near future (see Figure 1), companies have been rapidly cutting costs enterprise-wide to help protect future
margins. Studies have shown in the first year a drug comes off patent, revenue often decreases by more than 80%.

**FIGURE 1**
The 2008–2013 Patent Cliff for Major Blockbuster Drugs

<table>
<thead>
<tr>
<th>2008 Expirations</th>
<th>2010 Expirations</th>
<th>2012 Expirations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risperdal (Janssen)</td>
<td>Lipitor (Pfizer)</td>
<td>Singular (Merck)</td>
</tr>
<tr>
<td>$4.9B</td>
<td>$12.4B</td>
<td>$4.3B</td>
</tr>
<tr>
<td>Effexor XR (Wyeth)</td>
<td>Advair (GSK)</td>
<td>Lovenox (sanofi)</td>
</tr>
<tr>
<td>$3.8B</td>
<td>$7.7B</td>
<td>$3.3B</td>
</tr>
<tr>
<td>Fosamax (Merck)</td>
<td>Plavix (BMS)</td>
<td>Cozaar (Merck)</td>
</tr>
<tr>
<td>$3.0B</td>
<td>$5.6B</td>
<td>$3.6B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2009 Expirations</th>
<th>2011 Expirations</th>
<th>2013 Expirations +</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enbrel (Amgen)</td>
<td>Zyprexa (Lilly)</td>
<td>Rituxan 2014</td>
</tr>
<tr>
<td>$5.3B</td>
<td>$4.7B</td>
<td>$5.4B</td>
</tr>
<tr>
<td>Prevacid (Abbott)</td>
<td>Seroquel (AZ)</td>
<td>Nexium 2014</td>
</tr>
<tr>
<td>$3.3B</td>
<td>$4.5B</td>
<td>$5.2B</td>
</tr>
<tr>
<td>Lamictal (GSK)</td>
<td>Actos (Takeda)</td>
<td>Neulasta 2015</td>
</tr>
<tr>
<td>$1.7B</td>
<td>$2.9B</td>
<td>$4.7B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2013 Expirations +</th>
<th>2014</th>
<th>2017</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rituxan</td>
<td>$5.4B</td>
<td>Avastin</td>
<td>$4.0B</td>
</tr>
<tr>
<td>Nexium</td>
<td>$5.2B</td>
<td>Remicade</td>
<td>$3.7B</td>
</tr>
<tr>
<td>Neulasta</td>
<td>$4.7B</td>
<td>Epogen</td>
<td>$2.5B</td>
</tr>
<tr>
<td>Abilify</td>
<td>$2.2B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: IDC Health Insights, 2010

These patent concerns, combined with shrinking product pipelines, have created an environment where mergers and acquisitions have become commonplace as organizations try to both bolster and diversify their product portfolios to stabilize and grow near-term revenue.

Widespread outsourcing of anything not considered strategically core to the business has also become the norm to help cut costs. Clinical trials, manufacturing, supply chain management, and even some aspects of sales and marketing are now all strong candidates for outsourcing consideration. However, while purchasing pipeline and leaning out the organization can help massage margins by improving the top and bottom lines in the near term, the industry realizes that long-term sustainability is dependent on a greater evolutionary change taking place.

Companies are beginning to shift their R&D approach from the traditional high-volume broad population–based blockbuster model to a model that going forward will focus more sternly on driving higher-profit lower-volume solutions for particular subpopulations and individuals. Advances in bioinformatics and genomics have led to the burgeoning field of pharmacogenomics, which focuses on optimizing
drug therapy based on an individual's genetic variations. These advances, combined with the general movement toward personalized medicine, have become a field that holds great promise for the life sciences and for healthcare overall. There are also strong expectations for further advancing personalized medicine through knowledge gained from proteomics, metabolomics, and other cutting-edge areas of research such as current studies of the human microbiome. While these areas of research are promising and represent the next phase of industry evolution for the life sciences, they do not come without challenges, many of which are technology related.

From an IT perspective, companies are undergoing significant application portfolio rationalization, software consolidation, and IT centralization. This is being done to reduce redundancy across divisions; consolidate decision making to capture economies of scale; reduce development, maintenance, and support costs for major systems; and, most importantly, eliminate data silos to improve cross-department integration. While these efforts represent foundational changes underway, one of the largest challenges for life science companies going forward will be their ability to aptly manage and harness the vast amounts of big data generated across the business that continue to grow exponentially. Companies' ability to effectively navigate this digital ocean of information across all areas of the business will ultimately determine the success of both IT sustainability and broader business strategy.

**Research and Development**

Research and development efforts are quickly being inundated with new types and volumes of data that the industry fears could rapidly overwhelm current IT infrastructure and paralyze organizations. While whole genome sequencing (WGS) data is regularly mentioned as a likely suspect, scientific data sources (i.e., proteomic and metabolomic data) and data from across the broader health ecosystem (i.e., patient clinical data, DICOM imaging, and HEOR data) are also set to contribute to this data conundrum. Medical imaging data in new drug submissions is also becoming an increasing challenge. Even longstanding clear-cut tasks in R&D like searching patents and scientific journals to find and potentially extract drawings, figures, and articles on chemical compounds are becoming an increasingly challenging search and extraction effort across tens of millions of records.

With the cost of WGS now approaching $1,000, we are entering an age where WGS technology will soon be broadly available both to researchers and in the clinic. Technological advances leading to the commoditization of WGS will enable providers to routinely capture patient WGS data as part of disease management and allow disease researchers to regularly sequence large disease tissue repositories,
enabling rapid relationship discoveries between genetic variation and disease manifestation. Managing and analyzing these large sets of data will grow increasingly dependent on cloud solutions, which are already beginning to emerge in the industry.

Both discovery research and clinical development beg the need for enhanced cloud computing, storage, and actionable analytic capabilities, but in very different ways. Discovery research has no regulatory constraints limiting data access, so in many cases (when data security concerns are not high), commercial public cloud services are well suited and have already been heavily exploited. However, clinical development data requires robust security because of privacy and trial control concerns, which makes private clouds an essential requirement. Examples include management of electronic laboratory notebook data in discovery research and retrospective analysis of archived clinical data sets in clinical development. Both are areas where data volumes are already beginning to reach the tipping point.

Life science companies also need to work more effectively with external partners going forward, such as collaborations with biotechs, medical centers, hospitals, and other healthcare payers and providers. For these and other more conventional external relationships (e.g., BPO, CRO, and other outsourcing services), it will be critical that data, information, and knowledge are shared in ways that are secure, real time, regulatory compliant, and easily searchable and that enable improved strategic and tactical analysis at the point of decision making.

Data agility will be central to facilitating much of the anticipated progress and developmental breakthroughs on which life science companies are staking their future. As a result, it is important that companies eliminate any remaining data silos as part of a shift in researcher mindset from "me" to "us" across the organization. This will facilitate simple yet innovative advances around data utilization such as enabling "Google like" searches on organization-wide data resources and the ability to create data-initiated automated alerts to routinely support both operational and strategic decision making.

**Manufacturing and Supply Chain**

The explosion of big data within pharmaceutical manufacturing and distribution is still in its infancy and only now just beginning to creep up, but over the next few years as item-level serialization takes hold, the volume of data that the industry will need to manage and analyze in real time is set to easily surpass petabytes of transactional information.

Industry pressure to adopt item-level serialization within the life sciences has steadily increased over the past several years as the
presence of counterfeit drugs has grown and instances of drug theft (stolen delivery trucks, burglarized warehouses, etc.) have become more commonplace. Although instances of counterfeit drugs often make the news in the United States, in many developing nations, their presence is an everyday reality. As many as one in five deaths from malaria each year are suspected to be related to counterfeit drugs that provided no therapeutic value. And in Africa particularly, counterfeit medication is suspected to be responsible for hundreds of thousands of deaths annually, according to the World Health Organization. The ability to scour billions of transaction records in real time to identify patterns that resemble counterfeit activity could ultimately lead to significant improvements in world health by allowing companies and authorities to respond to and dispatch resources as necessary in a more timely fashion.

For these reasons, increased supply chain visibility and security are top of mind among most pharmaceutical manufacturers. Companies still lack visibility to the route in which their drugs travel from distribution to the end consumer. Similarly, in the event of a drug recall, locating all products and ensuring complete removal from the market are significant challenges that often occur with little efficiency in execution.

While legislation for drug pedigrees has been delayed, forward-looking companies are still pushing ahead with efforts to serialize since beyond enabling track and trace there are many other benefits to be gained, both operationally and strategically. This includes areas such as channel optimization, inventory reduction, increased perfect order performance, better forecasting and production planning, streamlined drug recalls, improved supply chain analytics, enhanced contract management, and reduced revenue leakage from errors in product returns, chargebacks, rebates, and concealed shortages. As IDC Health Insights has previously documented (see Business Strategy: Revenue Leakage — Pharma’s $11 Billion Problem, IDC Health Insights #HI220793, December 2009), these areas alone hold more than $11 billion of potential uplift to the industry.

As item-level serialization expands, big data will quickly become a big problem if manufacturers don’t proactively invest in building out the necessary infrastructure and processing power to ingest all of the transactional information and search it effectively in real time as well as drive actionable insights from their analysis.

**Sales and Marketing**

The pharmaceutical sales and marketing landscape is undergoing significant transformation because of increasing regulatory pressure (Physician Payments Sunshine Act, state-level aggregate spend regulations, gift ban laws, etc.) and the broadening impact of U.S.
healthcare reform. Physician availability is dwindling, consumerism in healthcare is growing, and payer influence throughout the healthcare system is strengthening. These broad changes are creating numerous challenges to traditional go-to-market approaches, leaving pharmaceutical sales and marketing organizations starving for new ways to measure and monitor promotional effectiveness, optimize campaign management, and improve overall sales and marketing intelligence.

Historically, achieving these goals has been difficult because of the siloed nature of market data residing across the organization and the backward-looking "reporting based" nature of most analysis that organizations have conducted. However, the industry is now quickly approaching a valuable opportunity window that should help this to change. Aggregated spend compliance is now forcing pharmaceutical companies to carry out massive integration and data aggregation exercises that will ultimately bring together data spanning the entire enterprise. As companies complete these initiatives, despite the recent aggregate spend compliance deadline extension granted by CMS, for the first time in history, companies will now have a single 360-degree view of every company touch point with all healthcare practitioners across all brands in U.S. geographies. And with similar regulations lurking across Europe, this newfound "aggregate view" will continue becoming more global. The availability of this new visibility will present a unique opportunity for pharma's to radically improve their sales and marketing analytical capabilities; however, it will also further amplify already mounting big data challenges.

Aggregate spend data will only be one of several data sets that are excessive in size. Prescription data, along with various forms of marketing information (digital media assets, social media data, etc.) are already reaching their choking point. Many pharma's are only harnessing a small fraction of the value contained in Rx data, and social media analytics is only now just beginning to approach its tipping point, which is set to drive data challenges surrounding the ingestion, storage, and analysis of voluminous online chatter.

There has also long been a fundamental disconnect between sales and marketing and the revenue management side of pharmaceutical manufacturing. Try asking a pharmaceutical company to tell you which U.S. state is the company's most profitable. Many companies simply could not answer the question or would take months to piece an answer together that accurately captured all factors affecting profit margin by state. This is because of poor data organization (i.e., several pieces needed for this puzzle live separately across silos), the enormous volume of data involved in this analysis, and many companies' inability to properly search and analyze all of the relevant information. Breaking down these walls by pulling channel information, contract data, rebate claims, and other strategic data sets across the organization to be correlated against Rx information and
impending aggregate spend data will improve sales and marketing decision making. It will also help organizations better understand the connections between various sales and marketing drivers (base price, discounted price, promotional activity, etc.) with the resulting effect they have on revenue by channel. This increase in revenue intelligence will help companies more accurately answer many strategic questions, such as:

- Why are products missing forecasts in some regions and not others, and how do we fix this problem?
- How does additional spending by channel generate incremental revenue by brand and region?
- Are channel initiatives actually delivering on their expected ROI?

This insight could also help companies in Europe better understand the margin effect of parallel trade and enhance their ability to utilize promotional activity as a countering force. But regardless of region, better cross-utilization of market data should prove very valuable, leading to increased market insight, more proactive decision making, and a potential competitive advantage for early adopters.

**FUTURE OUTLOOK**

With life science organizations' propensity for saving all information virtually forever, it is critical to decide how this data will be stored and managed. Data will live longer than any application or piece of infrastructure. Data volume will grow to make storage tiering an economic necessity as there will be more data than can be analyzed at one time and the best platform for analysis will not likely be the best platform for bulk storage. Data needs to be able to move back and forth between high-performance analytic environments and highly efficient massive scale content repositories many times over its life cycle as even archived projects need to be considered in ongoing and future analysis. Data growth will also make data migrations and traditional protection practices become untenable. Organizations will need dynamic, virtualized storage infrastructures with the ability to separate data from the applications that created the content. This will enable easier movement of data among storage tiers, secure sharing of information across multiple parties, and the ability to repurpose data for new applications.

Big data is not just about storage. The volume of data is only a small piece of the larger puzzle. Rather, the crux of big data initiatives is to allow life science organizations to produce strategic and scientific value from their big data through better utilization and analysis of information. In R&D, this will help drive more scientific discoveries, better global collaboration, reduced cycle times, and more efficient
FDA submissions. On the commercial side, optimizing supply chain, fighting counterfeit drugs, and improving multichannel marketing as well as the creation of a more agile and responsive patient-centric sales and marketing organization are all to be expected. There is also more interaction occurring now between R&D and the commercial side of the business (earlier communication between manufacturing and clinical development to prepare manufacturing and distribution strategy, clinical reporting of aggregate spend data, etc.). Even search and eDiscovery are becoming more of a big data challenge as most information is now stored electronically in different silos built around applications, lines of business, or geography. Finding the right content in this growing disparate set of information has become an extremely time-consuming and costly endeavor.

Many areas across the enterprise will both benefit from and be challenged by big data. However, one thing is clear: Challenges surrounding big data will be in focus for many years to come. The industry's biggest need is to have all of the ensuing big data efficiently stored, easily searchable, and available for comprehensive analysis, all in real time. Implementing appropriate infrastructure to ensure this happens remains a top priority for the life sciences. As companies approach these initiatives, important goals to keep in mind when approaching big data include:

- Create organization-wide consensus on content handling and data ownership to remove data access hurdles.

- Begin consolidating silos into dynamic storage infrastructure that will support the billions of objects and petabytes of data on the horizon.

- Seek out technologies that can search, tier, and automatically manage the life cycle of data across existing silos according to content and metadata.

- Focus on standardizing and automating data capture, storage, tiering practices, and management to enable better exploitation of data resources across the enterprise.

- Proactively build effective data sharing with customers, external vendors, and partners into your data strategy.

- Ensure appropriate data security exists from both an organizational and an IT standpoint.
ABOUT IDC HEALTH INSIGHTS

IDC Health Insights provides research-based advisory and consulting services that enable healthcare and life science executives to:

- Maximize the business value of their technology investments
- Minimize technology risk through accurate planning
- Benchmark themselves against industry peers
- Adopt industry best practices for business/technology alignment
- Make more informed technology decisions
- Drive technology-enabled business innovation

IDC Health Insights provides full coverage of the health industry value chain and closely follows the payer, provider, and life science segments. Its particular focus is on developing and employing strategies that leverage IT investments to maximize organizational performance. Staffed by senior analysts with significant technology experience in the health care industry, IDC Health Insights provides a portfolio of offerings that are relevant to both IT and business needs.

Copyright Notice

Copyright 2012 IDC Health Insights. Reproduction without written permission is completely forbidden. External Publication of IDC Health Insights Information and Data: Any IDC Health Insights information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Health Insights Vice President. A draft of the proposed document should accompany any such request. IDC Health Insights reserves the right to deny approval of external usage for any reason.