ELECTRONIC LAB NOTEBOOKS
HELPING ORGANIZATIONS BOLDLY GO
BIOVIA WORKBOOK CASE STUDY[1]

Forty-two years ago, Arthur C. Clarke envisioned the HAL 9000, a sentient computer capable of directing a space mission to Jupiter. At the Symyx Symposium in Barcelona(2010), Stephan Taylor, director of project and process optimization systems in process R&D at Bristol-Myers Squibb, envisioned another version of HAL: the Highly Automated Lab, managed by a sentient electronic lab notebook (ELN).

1. This case study summarizes a number of talks presented at the 2010 Symyx Symposium with reference to Symyx Notebook. Symyx was acquired by Accelrys in 2010, and the ELN was renamed Accelrys Electronic Lab Notebook. Accelrys was acquired by Dassault Systèmes in 2014, and the ELN was renamed BIOVIA Workbook.

Challenge:
Increased volume of life science information; proliferating data silos; new research techniques, product delivery methods and regulatory requirements; "homo sapiens-based" data integration

Solution:
Implement next-generation laboratory informatics with BIOVIA Workbook (previously Accelrys Electronic Lab Notebook, formerly Symyx Notebook).

Benefits:
- Improved data quality
- Better informed, higher quality science
- Standardized analytical methods
- Faster process execution
- Improved efficiency and productivity

Like the HAL in 2001: A Space Odyssey, Taylor's HAL can advise scientists about potential problems in their "mission objectives"—in this case, launching a new experiment in an ELN based on a previously input plan. "Dave, you have not taken sufficient precautions to prevent a thermal runaway. The heat capacity of the solvent and the available cooling capacity are insufficient to maintain a safe reaction temperature."

The reference may be science fiction, but Taylor considers ELNs a key component of an electronic lab environment that helps scientists execute faster, more informed, high-quality science. It's a vision scientists have talked about for decades. But Taylor's colleague Jason Bronfeld, executive director of preclinical and pharmaceutical development informatics at BMS, noted that the technology is finally in place to support what he calls the "information management journey" on which research organizations must now embark.

"The transition from paper-based systems to electronic ones is the most productive transition we will make as an industry," said Bronfeld. "But it's just the first step on an information management journey that will move us from reactionary, artifact-centric, locally optimized informatics to strategy-driven, process-centric, globally optimized environments where we aren't thinking about the things the process produces, but instead about the process itself and making it leaner and more efficient."

ONGOING CHALLENGES

On the surface, the informatics challenges currently plaguing life science R&D seem perennial: too much data, siloed in individual departments where it is inaccessible to drive most research decisions (see Figure 1) . These problems have been compounded in recent years by unprecedented technological and business pressures, and these pressures will only increase in the next decade.

For instance, merger and acquisition activity has always been prevalent among life science organizations. Recently, though, many companies have turned to outsourcing and contract research to stimulate pipelines. This "externalization" of research has changed what it means to "own" an asset, according to Ashley George, director of the strategic IT portfolio for discovery at GlaxoSmithKline. In what he called "the old days," organizations owned all their assets and followed a fully internal model with a linear workflow from design to synthesis to reporting. Today, George explained that the true business problem is externalization, which makes IP harder to trace. And as technology stacks change over time, organizations need to find a way to work across complicated layers of interaction to achieve the ultimate goals of designing, making, and testing compounds.

New research techniques, product delivery methods, and regulatory requirements that aim to make patients the true end game of any development effort also challenge organizations to rethink their approaches to R&D. Paul McKenzie, global head of pharmaceutical development and manufacturing sciences at Centocor Research & Development Inc., cited an influx of monoclonal antibody-based platforms, developed internally and in collaboration with partners, as an impetus to revamp development activities. Determining how best to capitalize on this platform requires new methods of making and delivering these new therapeutics. "If we're inefficient today making monoclonal antibodies and putting them in vials, imagine how we'll do sorting through new, diverse pipeline opportunities and novel delivery methods," said McKenzie.

Further complicating matters is the onslaught of data that a single pharmaceutical company and its partners can generate. The sheer volume of information to sort through during discovery, development, and manufacturing has always been daunting, but several speakers noted that the burden has finally become unsustainable. "The user is the informatics," said Taylor, citing the concept of the Renaissance Man. During the Renaissance, and even today, knowledge was transferred primarily in writings that people read and interpreted and

"The transition from paper-based systems to electronic ones is the most productive transition we will make as an industry."

— Jason Bronfeld, Executive Director,
Preclinical and Pharmaceutical Development Informatics
Bristol-Myers Squibb

tested experimentally. So it was conceivable that someone could know everything there was to know. That's just not possible given the volume of information produced today in modern discovery, which has meant that scientists have specialized in smaller and smaller areas and created more and more data silos.

McKenzie has another term for this unsustainable informatics model: Homo sapiens-based data integration. "We generate all these nice little folders, and we assemble them and carry them up the mountain for QA to look at," he described. "Eventually, we are left with piles of paper that we can't do anything with."

Coincidentally, it was the quintessential Renaissance Man—Leonardo Da Vinci—who is widely credited with creating the first lab notebooks. Even individuals who knew everything there was to know needed a place to record that knowledge. Today, ELNs function well in that role. But as McKenzie noted, the challenges faced in modern R&D can't be solved just by investing in a particular tool. "It's about how you fundamentally change the way you work and implement technology to make it work for you," he said.

Bronfeld put it another way. "Your ability to conceive paradigmatic change is limited by the tools you use to experience it," he said. "As one adage goes, it is tempting, if the only tool you have is a hammer, to treat everything you have as a nail. Today, discussions often get sidelined trying to optimize a local process in the weeds, when you actually need a step change that would render that local process completely obsolete."

Taylor summed the challenges up as follows. "When you look at being able to mine and look for relationships between things, those relationships are only as good as the data that you have," he said. "Historically, the earth was flat, and we were happy with that until the data told us it wasn't. Or think about how long we thought of the universe in terms of visible light because we couldn't see other wavelengths. The design space is limited by the tools we use to look at it, and as the tools improve, so will our ability to see what's possible."

STANDARDS: ONE WAY TO SEE BEYOND

ELNs certainly have the potential to free the design space by unlocking critical research discoveries from their paper prisons. But organizations with experience implementing ELNs have found that electronic systems can sometimes reveal more fundamental problems in how organizations describe and conceive of data. "The only similarity among our many development sites is that they use a lab notebook," said McKenzie. "How they use materials, name them, write them down, and put data together is all completely different." Such problems are not isolated to paper-based systems—those development sites could all use an ELN, but if that was still the only commonality, data management would be no easier than with a paper-based system.

Vocabulary problems occur even though many research groups essentially do the same types of tasks. "You cannot imagine how many of the same variables are named 20 different ways," McKenzie commented. The variability exponentially increases when you include external sites—even though most life science companies are all doing the same things.

Take sourcing. Scientists looking for a reagent want to know where it is, whether their company owns it, and, if not, where

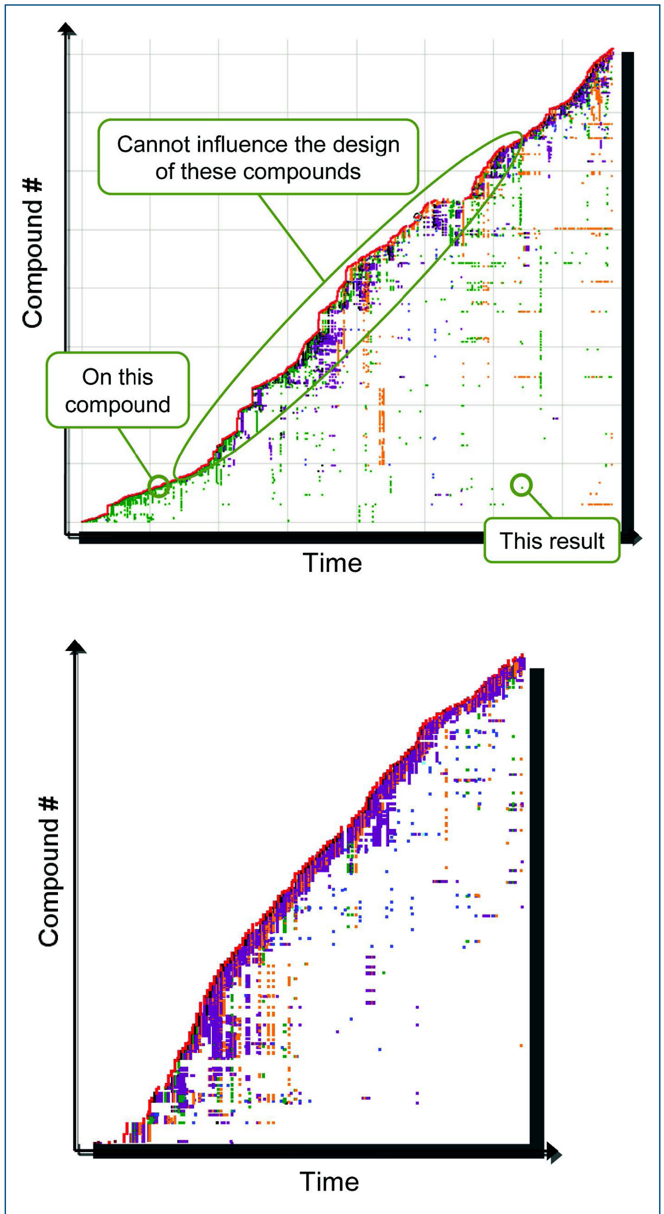

Figure 1: Event-based analysis deployed by one research organization to map out all the data collected on compounds over time. The red line marks when a compound was first registered; colored dots to the right track the discovery of compound details. The first chart above indicates how information was often obtained too late to inform work on circled compounds (thin line because information comes in over long time). The second chart above shows how informatics implementation tightened the time between discovering a compound and capturing additional details (thick line indicating faster knowledge capture) with the shorter cycle time enabling better informed decisions.

they can buy it. The questions are simple and universal to all organizations, yet every life science R&D organization struggles separately with ways to provide this capability to scientists—and may even offer separate sourcing solutions to different groups. Why? "Sourcing isn't competitive, it's just something we need to do," said George.

McKenzie noted similar problems for common tasks in process development, analytical development, process manufacturing, and quality control. "I get no proprietary advantage in running a centrifuge, unless I can run that important compound on it tomorrow rather than next year," he said. Yet organizations continue to spend enormous effort validating instruments and methods and finding ways to share methods and templates with colleagues in other disciplines.

McKenzie proposed that the industry could leverage ELNs to establish a system-independent recipe model to harmonize data exchange. His vision is to configure reusable libraries of standard unit operations that could then be exchanged between ELNs and execution systems in manufacturing and research. McKenzie likened these recipes to those used in cooking. All recipes, whether baking a cake or running a K*f*, consist of a procedure, equipment requirements, and a formula containing raw materials, raw material quantities, and process parameters (temperature, time, etc.) (see Figure 2). An agreed-on, common vocabulary for materials and recipe steps associated with common analytical and process methods would allow instantaneous transfer of methods between teams in an organization. Moreover, organizations could spend less time on technology transfer and validation and focus instead on managing individual parameters, which is ultimately where the real science is done.

Standards efforts are already underway, with one of the most high profile being the projects headed up by The Pistoia Alliance (see last page). And BIOVIA is already spearheading development of an analytical method standard in collaboration with several companies, including BMS. It is tempting to dismiss vendor-led standards initiatives as attempts to set prescriptive workflows that only can be managed with the vendor's product. But Bronfeld noted that such perceptions represent a myopic view of vendor/customer relationships.

"When you get past the bag o' applications that we all have and think about how to thread them together, your process pops out," said Bronfeld. He explained that independent standards ultimately benefit everyone, but creating them requires a true partnership between vendors and research organizations. Research organizations know the domain requirements, which vendors need in order to develop the best tools. Conversely, vendors have the informatics skillset to build appropriate tools, which research organizations need to advance on the informatics journey.

"If I have a good partner building good tools, we succeed

Dark Chocolate Cake

Formula				
Raw Materials	Raw Material Quantities	Process Parameters	Recipe Procedure (Process Description)	Equipment Requirements
All-purpose flour	2 cups	350 °F	Preheat oven	Oven
White sugar	2 cups		Grease and flour a pan	9x13 inch pan
Unsweetened cocoa	3/4 cup		In a bowl, combine the flour, sugar, cocoa, baking soda, baking powder, and salt	Bowl, spoon
Baking soda	2 teaspoons			
Baking powder	1 teaspoon		Make a well in the center	
Salt	1/2 teaspoon		Pour in the eggs, coffee, milk, oil, and vinegar	
Eggs	2	3 minutes		
Cold brewed coffee	1 cup		Mix until smooth, batter will be thin	
Milk	1 cup		Pour into prepared pan	
Vegetable oil	1/2 cup	35 to 45 minutes	Bake in the preheated oven	
Vinegar	2 teaspoons	72 °F	Allow to cool	Cooling rack

Figure 2: Centocor's Paul McKenzie likened process development to baking. Both require a recipe containing procedures, equipment requirements, and a formula that includes raw materials, raw material quantities, and process parameters such as time and temperature. The challenge in process development is creating a common vocabulary to describe materials and steps.

together," Bronfeld said. "The goal is to share the journey, because we're all trying to get to the same place. If you're arguing about what features will be in rev 6.5.2, well, you have the wrong partner."

THE INFORMATICS JOURNEY

The journey Bronfeld urged vendors and life science organizations to share begins where most companies are today: at a crossroads between paper-based and electronic systems (see Figure 3). Some companies have simply transitioned to a "paper-on-glass" metaphor that essentially reproduces paper artifacts electronically. This important first step, combined with identifying process and data standards such as those recommended by McKenzie, propels organizations down a path that creates clean, tractable data, drives out human variability, and increases data integrity.

"The goal of the early informatics journey is to become fully electronic, where you never have to exit out of a system to record what you are doing," said Bronfeld. He and several other speakers referred to "self-documenting processes." Today, many workflows require scientists to step away from work in progress to explain what they are doing. Such activities become obsolete in a fully electronic world. "Think about compound registration," said Bronfeld. "It's a non-value added task. Why can't a compound just be registered? As organizations proceed on the informatics journey, data sets get richer and enable execution by first eliminating the need for scientists to document that they are executing."

Taylor pointed out that after so many years being overwhelmed by data, organizations can find it counterintuitive to consider the power associated with collecting more raw data. But he explained that this is because most organizations still expect the users to function as the informatics. "If we do a good job collecting data and providing metadata and context, we end up with a giant, multivariate analysis problem, and this is what computers do best," Taylor said. "We should eventually be able to point computers at the problem and get the computers to sort out the relationships."

This leads to the predictive power organizations have long craved. "At the end of the journey, you achieve understanding-based execution where you improve quality, drive out variability, and leverage the computer to execute," said Bronfeld. Taylor gave a specific example. "The sentient ELN of 2025 will actually understand what we want to know because it will know all the materials and chemistries we've done," Taylor said. "It should be able to not just make suggestions, but actually tell you the overall pros and cons of, say, a particularly planned synthetic route. This means you can quickly evaluate methods before doing any experimentation."

WORKING IN TODAY'S AUTOMATED LAB

Unfortunately, the most frustrating aspect of the informatics journey is that it can't be planned or prescribed. Bronfeld was adamant that true, step-change innovation comes from being broad, not deep, and never happens when organizations set out to create something entirely new.

"Paradigm stage innovation is highly unforeseeable," said Bronfeld. "The best thing to do is to know your destination, be open to the punctuated equilibria of IT development, and take notice of the places where ambition collides with reality. From

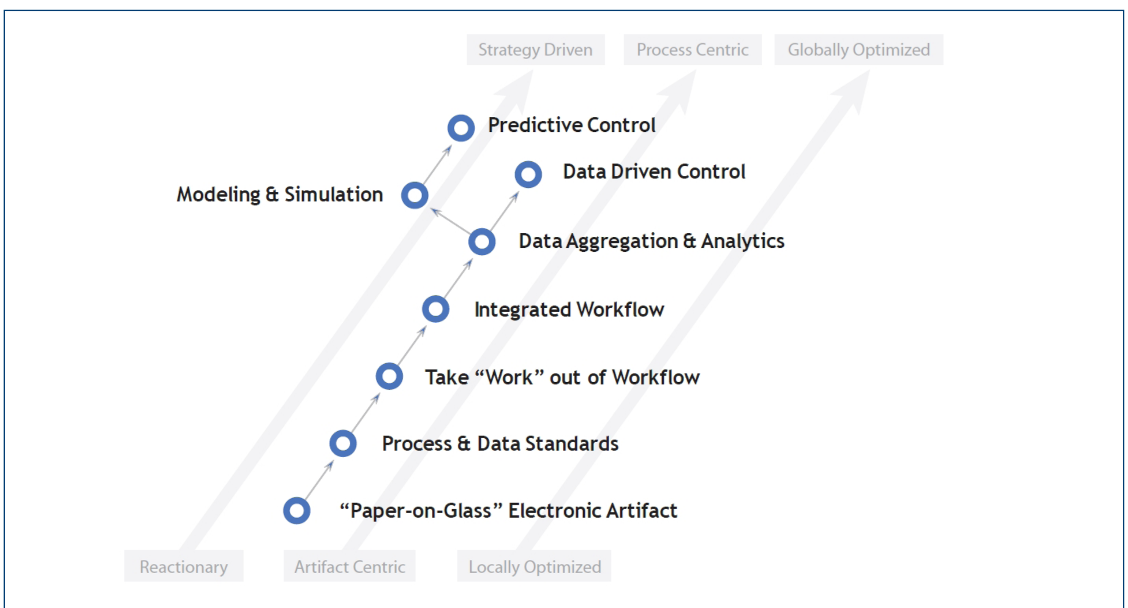

Figure 3: The steps along the information management journey enable organizations to move from reactionary, artifact-centric, locally optimized informatics to strategy-driven, process-centric, globally optimized environments

THE PISTOIA ALLIANCE: PUTTING THE INDUSTRY IN INDUSTRY STANDARDS

Originally conceived in 2007 by a group of informatics executives from four major pharmaceutical companies, the Pistoia Alliance has transitioned from a grassroots effort to a project-centered organization set on providing several key concrete standards to facilitate life science research.

In 2010, Pistoia established three working groups focusing on the following projects:

- ELN query services aims to develop standards that can be used across different notebooks serving different domains (initial attention on chemistry data types). BIOVIA has taken an active role in this working group.
- Sequence services is exploring whether a hosted service for storing sequence data would appropriately cater to security and scalability needs in life science.
- SESL will demonstrate the feasibility of an open knowledge-brokering framework standard that will reduce the costs of integrating disparate data sources. This working group is funded as a partnership between Pistoia and the European Bioinformatics Institute.

"We've kept Pistoia's initial program deliberately small focusing on only a few projects," said Nick Lynch, Pistoia's president. "We don't want to overextend and end up providing little of use."

The cloud is a logical place to provide these standards once they are available. With everyone trying ultimately to do the same thing, the cloud offers a way to make services that many companies have tried to provide internally easily accessible to everyone—without forcing organizations to punch holes in their corporate firewalls.

Pistoia membership is open to individuals and organizations.

To get involved, visit http://www.pistoiaalliance.org/.

there, you can plot a strategy that moves you from where you are now to a near-term future state and, ultimately, further along the journey."

In the context of the journey, an ELN is a critical first step rather than a magic bullet. Implementing an ELN enables organizations to begin to centralize electronic data and identify legacy tasks and processes that are unnecessary in an electronic workflow. And with that electronic workflow established, the future state organizations desire is attainable. We know that, according to McKenzie, because other companies have achieved these levels of automation and control.

"Imagine us, several years from now, running 400 pharma plants from a 10x10 room," he said. "This is not science fiction. Many industries worldwide already do this. And they are doing fermentation, a reaction many of us also do. If they can do it, so can we."

About Dassault Systèmes BIOVIA

Dassault Systèmes BIOVIA provides a scientific collaborative environment for advanced biological, chemical and materials experiences. Sophisticated enterprise solutions from BIOVIA support Collaborative Science, Unified Laboratory Management, Process Production Operations and Quality/Regulatory Management, driving innovation for science- and process-based industries. For more information, visit http://www.3ds.com/products-services/biovia

Dassault Systèmes Corporate
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

BIOVIA Corporate Americas
BIOVIA
5005 Wateridge Vista Drive,
San Diego, CA
92121
USA

BIOVIA Corporate Europe
BIOVIA
334 Cambridge Science Park,
Cambridge CB4 0WN
England