

Common Knowledge

Two heads (or more) are better than one,
except when they don't share information.
That's where knowledge management comes in.

BY SALVATORE SALAMONE

What's more difficult than finding a needle in a haystack? Defining knowledge management.

Yet life science companies, collaborative software tool vendors, and industry consultants all seem to hold out great hope for this something they call knowledge management.

The reason is simple: With such massive data overload in the life sciences, companies realize it is to their advantage to find better ways to deal with and use this information. "We want to leverage biology knowledge for discovery," says Mark Osborne, assistant director of knowledge management at Millennium Pharmaceuticals Inc.

Capturing researchers' knowledge is vital to the success of an organization. "Your speed of innovating is determined by your experts' ability to integrate their new knowledge into a teachable framework, and your ability to mobilize it," says Victor Newman, chief learning officer at Pfizer Inc.'s internal Research University and author of "The Knowledge Activist's Handbook."

To use knowledge as a strategic advantage, life science companies take one of three broad approaches, which can be generically described as follows:

- Help scientists know what's already been done and by whom
- Extract information, relationships, or new insights from existing diverse data sources
- Capture the expertise and intellectual property of scientists

One approach in using knowledge management is to give researchers a means to better understand what's already been done and who did it. This is not a trivial point: "A common problem in pharma companies is that [researchers] don't know what's been done before," Osborne says.



Basking in the know: Mark Osborne, assistant director of knowledge management at Millennium Pharmaceuticals Inc., says researchers need to look also at the science being done outside company walls.

Hence a number of knowledge management projects aim to provide scientists and managers with tools that help them stay informed about what is going on within the company. This helps prevent researchers from duplicating a completed experiment or following a nonproductive path that has already been rejected. The worst-case scenario — one that happens far too frequently — is for a researcher to spend six months eliminating a particular drug candidate, only to find that a colleague had already drawn the same conclusion.

"It's not a matter of putting in the time to eliminate candidates; that's the normal procedure in any R&D project's early stages," says a manager at a Massachusetts-based pharmaceutical company who does not wish to be identified. "The frustration comes when you don't know some other group has already done the same experiment — it's a waste of resources, and that time is completely lost."

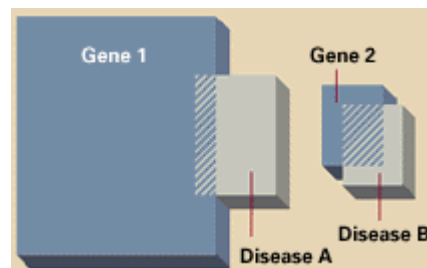
Informing researchers that a co-worker is interested in the same topic is one way to avoid such duplication. Such an approach can also reap other benefits — namely, time savings. For instance, scientists in the Drug Innovation and Approval (DI&A) group within Aventis Pharmaceuticals Inc. were able to save weeks of effort in certain projects by using a knowledge management system that quickly identifies internal experts in any given field.

At Aventis, a scientist based in Frankfurt, Germany, who was studying thrombotic and joint diseases began a project to isolate and culture macrophages. Using an in-house system called KMail, the scientist was able to find two colleagues in the company's Bridgewater, N.J., facility who had experience with the key protocols. "One helped with culturing protocols; the other helped with information on magnetic cell sorting," says Steve Sorensen, head of knowledge management systems for Aventis' DI&A group. "Using KMail cut about four weeks out of this one research [project]."

To set up the KMail system, Aventis used the knowledge management consulting firm Plural Inc. to help design and deploy the pilot project. This involved 435 scientists from Aventis labs around the world. The heart of the system is a collaborative product called KnowledgeMail from Tacit Knowledge Systems Inc.

Worth Weighting

A straight search of scientific literature for relationships between genes and diseases may give misleading results if only a raw number of co-citations is used as a gauge. The Harvard Medical School's Institute of Proteomics MedGene search program allows researchers to use different statistical weighting algorithms to help find relationships between less-often cited genes and diseases — relationships that might otherwise be obscured in comparisons that look at only the raw numbers of citations.



In this diagram, the area of each rectangle represents the number of citations of a gene or disease, and the hashed area represents the co-citations between a gene and disease. Even though both examples have about the same number of co-citations (the hashed areas are equal in size), it is clear that there is a tighter relationship between Gene 2 and Disease B than between Gene 1 and Disease A.

Moving KM to a Higher Level

A different type of knowledge management involves extracting useful information from diverse data sources. For example, researchers investigating a new drug candidate typically have to work with an assortment of proprietary chemical libraries, databases (both in-house and annotated versions of public databases such as GenBank), and internal documents. In addition, it is essential for researchers to monitor a variety of other sources, including scientific journals and competitors' and government and regulatory agencies' Web sites.

The challenge for life science managers is how to make better use of this information. "There's a vast majority of science not being done inside our walls — we need to relate that work to what's being done internally," Millennium's Osborne says.

Large global pharmaceutical companies spend, on average, about \$8.28 million annually on information content, according to Outsell Inc., a research consultancy that focuses on the information content industry. Outsell derived this figure from a survey of 12 top pharmaceutical companies (with average revenues greater than \$10 billion).

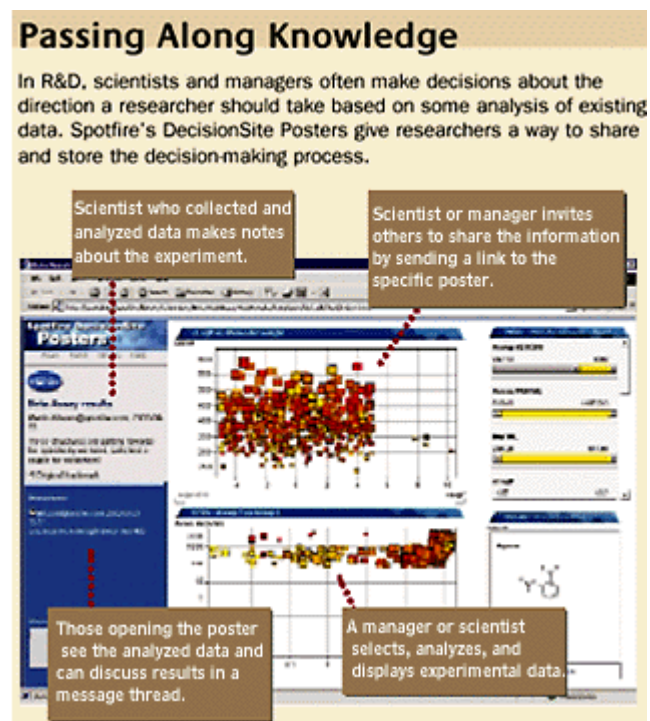
Not knowing exactly where to look, what to look for, or how to efficiently search can easily lead people astray. As one researcher put it during a break at a recent knowledge management seminar, "When I start a project, I don't know what I don't know. I usually ask all the wrong questions."

Consider that a typical Google search on a new topic can easily add up to hours — and that's just for a search confined to pages on the Internet.

So, time is the demon. How much time is spent searching databases and other sources for information? Outsell claims workers in all industries spend, on average, 25 percent of their time — 9.25 hours per week, to be precise — obtaining, reviewing, and analyzing information. The problem in life science companies is probably even worse, given the variety of information sources available. "Professionals in the [life sciences] spend 30 percent of their time looking for information," says Andy Michuda, CEO of the knowledge management and business process management consulting company Sopheon.

To make matters worse, with the hunt-and-peck search method that is commonly used, a scientist will not likely find hidden or less-than-obvious relationships between things (e.g., a disease-gene relationship), except through serendipity.

One way to handle this information glut is to simply use smart people with the right tools to do the work. Naturally, the difficult part is having enough qualified people to do the searching. Some companies are turning to co-sourcing — bringing in knowledge management consulting companies that supply the search tools and the subject matter experts to help answer questions, and set up the processes for submitting queries.



Five years ago, the Clinical Information Group at SmithKline Beecham (now GlaxoSmithKline) was expanding its role in support of clinical R&D. The group's six staffers had strong information management and science backgrounds, and it was being asked to deliver more analysis and interpretation of experimental results.

Recognizing that the group could use twice the staff, the company turned to Sopheon, a firm that provides software, consulting services, and experts to help companies make business decisions. With SmithKline Beecham, Sopheon researchers with expertise in chemistry, life sciences, and business and legal fields worked with the internal information scientists.

The relationship continues today, and GSK has developed metrics that gauge the effectiveness of the external experts. Among those metrics are response time, relevancy of the information delivered, and amount of editing required to meet the end users' requirements. According to GSK officials, the external information experts meet expectations more than 90 percent of the time.

What's in a Name?

A somewhat different approach to finding answers in numerous databases of unstructured data is to use more intelligent search tools. More intelligent tools are needed because common text search tools that find things so well on the Web do not work as well with life science databases, largely because there is no universal, consistent way to denote genes, disease names, or other entities in the scientific literature.

The problem with gene nomenclature is a notorious example. Different researchers and publications may adopt subtle variations in the name of a given gene. For instance, human genes are typically designated by upper-case symbols with no hyphens, such as G protein-coupled receptor genes GPR1, GPR2, GPR3, and so on. But some authors may refer to these genes as GPR-1, GPR-2, and GPR-3 or insist on using completely different lab names, perhaps in an effort to indicate precedence in the discovery of the gene. Whatever the reason, literature searches can typically miss important articles simply because of differences in gene abbreviations.

One way around this problem is to use tools that have life science-specific glossaries that take these variations into account. Last year at the 21st Century Life Science Technology Revolution conference, Joshua LaBaer, director of Harvard Medical School's Institute of Proteomics, addressed this issue: "The challenge is getting researchers to share their nomenclatures. There's an old saying that researchers would rather share their underwear than their nomenclature."

Researchers in LaBaer's institute have developed MedGene, an online tool that helps identify well-known and perhaps lesser-known relationships between genes and various diseases. MedGene summarizes biomedical literature and lets researchers detect what are called co-citation relationships between a disease and a gene. Essentially, MedGene searches the titles, abstracts, and medical subject main headings (the so-called MeSH indices) of more than 11 million Medline records. The MedGene application takes into account the various ways that different scientific publications denote the various diseases and genes.

Researchers using MedGene can find lists of genes associated with a particular human disease or diseases, and vice versa. They can use a handful of statistical methods to rank the importance of each of these lists, highlighting relationships that might otherwise go unnoticed.

For example, if no statistical weighting is applied and a simple mean number of citations is used, a commonly studied gene will get artificially ranked higher than a less popular gene (see "Worth Weighting").

Others take this a step further by combining statistical weighting with semantic analysis. Semantic analysis uses natural language processing technology to identify patterns between items within the text of an article. This combination of techniques was used by the team of data-mining company ClearForest and Celera Genomics to win the Association for Computing Machinery's Knowledge, Discovery, and Data Mining Cup competition last year. This contest required entrants to build a knowledge management system that first automatically curated thousands of articles on *Drosophila melanogaster* for FlyBase, the fruit fly genome database. The system used for the competition then had to note which articles had results of experiments on gene expression and which genes and proteins were involved in these experiments.

Getting Humans Involved

Knowledge management approaches based on statistical and semantic analysis can be useful in finding new links between genes and diseases from past experiments. But what would yield even more useful information would be the ability to understand why certain experiments were done or why a researcher made a decision to pursue one research path rather than another.

That is an entirely new level of knowledge management. Ideally, a company would like to be able to collect the intellectual property of its employees into a fully searchable database.

Capturing actual knowledge (as opposed to data) is not an easy task. Sociological problems must be overcome: Many researchers naturally balk at the idea of sharing hard-earned information.

"Researchers have been trained in university settings, and while they may have left academia for the corporate world, they have been imprinted with a research culture," says Bradford Kirkman-Liff, from Arizona State University's School of Health Administration and Policy. "Faculty members have a long tradition of exploring their own interests and pursuing their own agendas when it comes to research."

"Knowledge sharing is [typically] limited to traditional forums, such as journals or conferences," Kirkman-Liff says. "Ongoing, daily sharing of insights and techniques is limited. And the life sciences research culture does not promote organizational collegiality nor rapid dissemination of ideas."

Convincing employees to contribute to a common pool of knowledge is one thing. But how do you capture this knowledge?

Several products on the market give researchers ways to express their thoughts electronically so that they may later be searched (see "Passing Along Knowledge"). While many companies find these tools useful, some say corporate-designated people should ride herd over the process to ensure the right information is being captured.

For instance, Millennium Pharmaceuticals uses knowledge intermediaries (KIM). When a scientist creates a document, presents a paper, or gives a presentation, the KIM asks for the electronic version of that item and summarizes it. The KIM then passes this summary back to the original scientist to ensure the ideas are captured correctly. Once the two agree, the summary and the item are "published" to an internal searchable database called the Journal of Millennium Science.

This is part of a large knowledge management system at Millennium. From this one system, employees can pull together information scattered among multiple sources, including external Web sites, documents stored in the company's Documentum system, and internal databases (a sequence database, for example).

In this way, "we enable the exploration of external information and visibility to other Millennium findings," Osborne says.

Tying Everything Together

For years, companies in many industries have successfully managed structured data, using search tools to perform data mining of large databases to extract information that could be used for competitive advantage. For instance, financial service firms note spending habits of people in certain salary ranges or neighborhoods, allowing direct marketing campaigns or new services aimed at those who fit certain criteria. Similarly, experimental data can be analyzed and visualized to note trends and patterns that might accelerate a drug's progress.

The real challenge today is to manage unstructured data. The common approach has been to create databases with the charge of "Let's capture everything we do in case we need it again." Unfortunately, many companies have lost sight of the business purpose behind this process.

It's one thing to make progress on developing a new drug candidate; however, a researcher could spend all of his or her time learning everything about a single drug. But that's like operating in a vacuum. What if the FDA makes a pertinent regulatory decision, a similar drug from a competitor has problems in clinical trials, or lifestyle changes occur that might affect the drug's profitability?

That's why life science companies buy content in the form of subscriptions to science publications, patent libraries, and regulatory statutes. But searching all the relevant data with an eye on the business reasons for doing research is a change.

Life sciences is an area where knowledge management can play a vital role. Culling information from disparate data sources to make better business decisions is what knowledge management is all about. ●

ILLUSTRATION BY BRIAN CRONIN; PHOTO BY DANA SMITH

The bio-itworld.com web site is produced by Bio-IT World
500 Old Connecticut Path P.O. Box 9010 Framingham, MA. 01701

<http://www.bio-itworld.com/archive/031003/common.html>