



Bio•IT World

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

**FORWARD THIS ISSUE
TO A COLLEAGUE**



**SUBSCRIBE TO
BIO•IT WORLD MAGAZINE**



Welcome to your digital edition of *Bio•IT World* magazine

We hope the magazine is meeting your expectations. We encourage you to share this issue with your team to stimulate further conversations. [Share this issue.](#)

We are genuinely excited about the prospects for this field—and the magazine—in the coming year. We've cemented our position as the flagship publication of Cambridge Healthtech Institute (CHI), and will continue to play a prominent role in hosting and reporting on the best CHI conferences throughout the year.

We pride ourselves on publishing critical insights and analysis of innovations across the drug discovery pipeline — from molecular modeling of popular drug targets to biomarkers that discriminate cancer responders, and from data handling for next-generation sequencing to new strategies for increasing the speed and efficiency of clinical trials. We will work hard to surpass those stories in the coming year, continuing our pursuit of the most critical tools and strategies that epitomize the world of “predictive biology.”

We hope to continue to engage you with our editorial content, and within our network, and as always, we welcome any and all comments or suggestions —editor@healthtech.com

NAVIGATION KEY

- **Previous/Next Page:** mouse click (or keyboard) on left/right arrows to turn page
- **Contents:** takes you to the Table of Contents page
- **To Zoom:** In the browser version, simply click on the page to zoom in, and again to zoom out; in the enhanced PDF, use the zoom in/zoom out buttons in the top or bottom tool bar.
- **Return to Front Cover:** In the browser version, just click the double arrow button in the Nav bar; in the PDF, click on the Front Cover in the Tool bar.
- **Search:** Click on the search button to perform a full keyword or phrase search. You can search a single issue, or your complete archive.

Take the shortest path to the **RIGHT PATH.**



Symyx Isentris

The Freedom to Explore.

Out here, being able to make decisions quickly and confidently changes more than your perspective. It gets you to the right path faster. The same holds true in the lab. With Symyx Isentris® software, delays caused by technology and overwhelming amounts of data are a thing of the past. Instead you discover more, repeat less, and get it right the first time.

Symyx Isentris lets you search, browse, analyze, organize, and report on corporate, commercial, and local database information following a self-service model. You can share data, collaborate with colleagues, and gain valuable insights. The result is less time sorting data and more time exploring science.

*To learn more, visit
www.symyx.com/isentris*

*To attend a free webinar, go to
www.symyx.com/webinars*



© 2009 Symyx is a registered trademark of Symyx Technologies, Inc. All rights reserved.



Cambridge Healthtech Media Group

www.bio-itworld.com

Bio-IT World

Indispensable Technologies Driving Discov-

Trials

MAY | JUNE 2009 • VOL. 8, NO. 3



**MODELING AND
SIMULATION AT
NOVARTIS 32**

**EMBRACING THE
ELECTRONIC COMMON
TECHNICAL DOCUMENT 36**

**MICROSOFT'S AMALGA
LIFE SCIENCES 58**



Everybody
Loves
Chris

IT consultant Chris Dagdigan
kicks off the 2009 Bio-IT World Expo 8-18

WE DISCOVERED STEM CELLS. IT'S TIME TO DISCOVER US.

Ontario has been home to one breakthrough discovery after another. In 1961, Lasker Award-winning scientists James Edgar Till and Ernest Armstrong McCulloch proved the existence of stem cells. Earlier this year in Ontario, Dr. Andras Nagy and his team discovered a novel way to generate safer stem cells from somatic cells, bringing us closer to treating or curing a multitude of diseases, including Parkinson's, Alzheimer's, macular degeneration and autism using the patients' own cells. Other Ontario scientists have found novel ways to treat cancer with viruses and have made advances in many additional fields, including regenerative medicine and neuroscience. Dr. Tony Pawson won the 2008 Kyoto Prize for Basic Sciences for his work in cell signaling, becoming the first Canadian awarded this honour. We're committed to even greater achievements in the future. Isn't it time you made a major discovery of your own? Ontario. **The world works here.**



ONTARIO
CANADA

investinontario.com/research
1-800-819-8701



Discover us at BIO Atlanta
Booth #3311

 Ontario Paid for by the Government of Ontario.

**"How do we know
this lead molecule
is novel?"**



**“SciFinder—
of course.”**

Need to assess the novelty of substances?

SciFinder is the answer.

It includes CAS REGISTRY,SM the most comprehensive substance information available, integrated with relevant journal articles and patents.

Give your research team the highest quality and most timely scientific information resource.

Make SciFinder an essential part of your research process.

For more information about SciFinder, visit www.cas.org or e-mail help@cas.org.

an essential
SciFinder®—Part of the process.TM



SciFinder®

CAS is a division of the American Chemical Society

www.cas.org

Contents [05–06·09]



13



26



36

Feature

Learning to Embrace the eCTD

Sponsors, CROs, and industry providers learn the pros and cons of switching to the electronic common technical document. [36](#)

Up Front

8 The Dagdigian Doctrine

What's hot and what's too high on the "hype-meter" in bio-IT.

9 Eric Schadt Leads Network News

There are no linear pathways, only networks.

10 Will the Gene Microscope Change the World?

Clifford Reid predicts a paradigm shift in genome sequencing.

11 Responsibility in Personalized Genomics

Making genomics useful to the consumer is proving a challenge.

Bio-IT World Signature Series

Trends in Global Economic Development

Begins on page 21

21 Trends in Economic Development

22 Putting the Alpha in Alabama

Advancements in Clinical Trials

Begins on page 42

44 CROs Heed the Crisis Call

12 Perfect Ten

Bio-IT World announces the winners of the 2009 Best Practices Awards.

13 Bourne Free Access Nets Franklin Award

Philip Bourne wins the 2009 Benjamin Franklin Award.

14 Best of Show Winners

Recognizing the most innovative new products on the Expo floor.

17 The Vagaries of Genome Variation: Do You Copy?

Comprehending Copy Number Variation makes an impressive debut.

18 Mark Boguski's Search for Resounding Health

An online resource for intelligent disease management.

19 Antibody Docking on the Amazon Cloud

Inside the Box Cloud computing performs up to the hype.

20 Will Consumers Sustain Direct-to-Consumer Genomics?

Insights|Outlook Searching for the consumer genomics business model.

18 Briefs

Computational Biology

26 Hans Lehrach's Predictive Biology Philosophy

Alacris Pharmaceuticals applies modeling to individualized medicine.

29 Bioalma Launches novo|seek PubMed Search Tool

Intelligent search ranks documents according to relevance.

Computational Development

32 Novartis Savors Early Modeling Success

Under Don Stanski, Novartis' M&S franchise is paying dividends.

IT/Workflow

46 Information To Bank On

The UK Biobank hopes to manage a quarter of a billion data points.

48 Isilon's Data Storage Odyssey

CEO Sujal Patel hopes to remove impediments to progress.

Clinical Research

51 Social Networking Sites Embrace Clinical Trials

Patient recruitment is online and interactive.

54 Texting in Clinical Trials

Texting offers instant contact and feedback.

In Every Issue

5 What's on Life Sciences TV

First Base The Franklin Award winner hosts SciVee.tv. **BY KEVIN DAVIES**

58 Microsoft's Amalgam Life Sciences

The Russell Transcript Microsoft unveils the next on-ramp for their information highway. **BY JOHN RUSSELL**

6 Company Index

6 Advertiser Index

56 Educational Opportunities

First Base



What's on Life Sciences TV?

KEVIN DAVIES

The winner of the 2009 Benjamin Franklin Award, Australian Philip Bourne (see page 13), shares much in common with the previous winners of the Bioinformatics Organization's annual prize. He's a savvy computer scientist, journal editor, author, open access advocate and, of course, widely respected among his peers. But he may be the first to add TV mogul to his list of accomplishments. No-one's going to confuse Bourne with Rupert Murdoch, even though they share the same accent. But SciVee.tv, Bourne's multimedia venture for scientists, deserves rather more attention than it has garnered thus far.

SciVee.tv was launched by Bourne and Leo Chalupa at [UC Davis](#), and is backed by the [Public Library of Science](#), the [San Francisco](#) non-profit that has pioneered open access publishing

through excellent journals such as *PLoS Biology*. SciVee.tv is, in a glib sense, YouTube for scientists—indeed, Bourne unashamedly says the popularity of YouTube in his lab was his inspiration. It contains a broad selection of videos of scientists giving lectures, presenting posters (postercasts), summarizing their latest publications (pubcasts), and conduct-

Steve Mayo (Caltech) channels his inner Al Roker at SciVee.tv.

ing interviews. Some enterprising students are using the site to broadcast multimedia resumes.

The topics covered a range literally from agriculture to zoology, and the site comes with full Web 2.0 capability. Just like YouTube, videos can be bookmarked as favorites (or flagged). Comments are welcome, though few and far between, and communities and discussion groups are encouraged. Uploading is easy and the video navigational tools are improving.

The videos range greatly in quality, but even just a quick perusal of the site finds a lot of interesting material. For example, hot off the video recorder, there's [Harvard's George Church](#) speaking in March at a DOE Joint Genome Institute meeting on "Reading and Writing DNA." His lecture concludes with the

unpublished news that one of his Personal Genome Project volunteers is a carrier of a mutation for hypertrophic cardiomyopathy ([scivee.tv/node/10578](#)). Craig Venter is here also (taken from the same meeting), talking about institute's pioneering work in synthetic biology and metagenomics ([scivee.tv/node/10653](#)). Nobelist Sydney Brenner is recorded delivering a typically quirky two-part monologue on evolution and "what genomes can tell us about the past" ([scivee.tv/node/8449](#)).

Some of the videos are of impressive technical as well as scientific quality. Several researchers, including [Caltech](#) biophysicist Stephen Mayo ([scivee.tv/node/8463](#)) and UCSF geneticist Cynthia Kenyon, deliver slick powerpoint presentations on topics ranging from computational protein design to longevity genes in the nematode. But what makes them stand out is the use of green screen technology, just as TV weathercasters use, allowing the host to gesticulate in front of their presentation. And of course there's Bourne himself, offering a series of "ten simple rules" for getting published or, indeed, making a good presentation ([scivee.tv/pubcast/17500596](#)).

SciVee.tv and similar initiatives have a huge upside, disseminating information not only to active researchers and students but also helping to educate the next generation of scientists. SciVee categorizes videos for elementary and high-school children, although this needs a bit of work. One entry, "Cell isolation of flow compatible mouse CD4+ T cells" doesn't appear entirely appropriate for kids aged 2-12 years. It's also in Japanese!

SciVee is clearly popular in Bourne's home town of San Diego. For example, the site hosts the "nifty fifty" videos from the San Diego science festival. That popularity is not yet matched across the country or in other parts of the world, but Bourne's innovation deserves to succeed. We congratulate him on SciVee and his Benjamin Franklin award.

Life Science Webcasts

Here at *Bio-IT World*, we're increasingly enamored of the potential of video. As part of a Cambridge Healthtech Media Group initiative (see [www.bio-itworld.com/lsw](#)), we've been posting life science webcasts of some of our past Bio-IT World Expo keynote lectures and panel discussions, and plan to do so with this year's crop. We've also presented one-on-one interviews with thought leaders such as Phase Forward founder Paul Bleicher, [CollabRx](#) co-founder



Paul Bleicher sits down with Kevin Davies in a webcast

Marty Tenenbaum, and former director of the [Allen Institute of Brain Science](#), Mark Boguski, discussing his new Internet company, [Resounding Health](#) (see page 18).

Company Index

Abbott Laboratories	12
Accelrys	14
ActiveHealth Management	12
Acurian	51
Alacris Pharmaceuticals	26
Allen Institute of Brain Science	5
Alliance Health	52
Amazon	13, 18
Amgen	12, 27
Applied Biosystems	18
Apredica	44
Apyx	38
Aspera	14
Association of Clinical Research Organizations	44
AstraZeneca	12
Beijing Genome Institute	18
Bioalma	29, 14
BioTeam	8, 19
BioWisdom	12, 14
BlueArc	13, 14
Boston University Schools of Medicine and Public Health	11
Broad Institute	12, 13, 18
Caltech	5
Celera	26
Center for Information and Study on Clinical Research Participation	51
Children's Hospital of Philadelphia	12
CLC bio.	14
CollabRx	27, 5
Complete Genomics	10, 11
Covance	44
CycleComputing	14
David H. Murdoch Research Institute	58
Dell	18
DNASTAR	14
EMBL	17
ePharmaSolutions	12
e-SubmissionsSolutions.com	38
Exco iTouch	54
Fred Hutchinson Cancer Research Center	58
Genedata	12
Gene Network Sciences	9
Genentech	12
Genome Center at Washington University School of Medicine	18
Geospiza	18
GlaxoSmithKline	12, 38
GlobalSubmit	40
Good Products	14
Harvard Medical School	5, 11, 17, 26
HHMI	17
Hospital for Sick Children, Toronto	17
HP	15
Ignite Institute for Individualized Health	18
Illumina	18
Image Solutions	40
IMC	15
Infinite Peripherals	15
Indigo Bio Systems	15
InfoSys	15
Ingenuity	15
Inspire	51
IO Informatics	15
Isilon Systems	13, 16, 48
Knome	11
MediciGlobal	52
Merck	9, 12
Microsoft	9, 46, 58
National Center for Genome Resources	12
Navigenics	11, 18
NextBio	16
Novartis	51, 32
Ocanina Networks	13, 16
Octagon Research	36
Oracle	46
Panasas	18
ParaScale	16
PatientsLikeMe	51
PBS GridWorks	16
Phase Forward	16
PleaseTech	16
Plectix	16
PPD Inc.	44
ProSanos Corporation	12
Public Library of Science	5
QUOSA	16
Resounding Health	5, 18
Rosetta	9
Sage	9
Selvita	16
Sequana Therapeutics	26
Sophic Systems Alliance	16
Third Rock Ventures	11
TriLink Biotechnologies	16
UK Biobank	46
University of California, San Francisco	5
University of California, Davis	5,
University of California, San Diego	13,
University of Southern California	18
University of Trento Centre for Computational and Systems Biology	58
University of Washington	17, 19
Valiance	16
Vanderbilt University	12
Wyeth Research	12, 37
Xyntek	16

Advertiser Index

Advertiser	Page #	Advertiser	Page #	Advertiser	Page #
Adobe LiveCycle ES	42-43	Chemical Abstract Services	3	Insight Pharma Reports	53
www.adobe.com/products/livelcycle/		www.cas.org		www.InsightPharmaReports.com	
Barnett Educational Services	41, 49	CHI Publications	35	Ontario Ministry	Cover 2
www.barnettinternational.com		www.healthtech.com		www.investontario.com/research	
Bio Georgia	23	eCliniqua	34	Spain Biotechnology	24-25
www.georgiabiobiosciences.com		www.ecliniqua.com		www.spaintechnology.com/biotech	
Bio-IT World Conference and Expo Europe 2009	30-31	Educational Opportunities	56-57	Symyx	Cover 4
www.Bio-ITWorldExpoEurope.com		www.bio-itworld.com		www.symyx.com/isentrus	
Biomarker Bridge	55	ERT	45	Tessella	7
www.biomarkerbridge.com		www.ert.com/pro		www.tessella.com	
		GeneGo	Cover 3		
		www.genego.com			

This index is provided as an additional service. The publisher does not assume any liability for errors or omissions.

VOLUME 8, NO. 3

Editorial, Advertising, and Business Offices: 250 First Avenue, Suite 300, Needham, MA 02494; (781) 972-5400

Bio-IT World (ISSN 1538-5728) is published bi-monthly by Cambridge Bio Collaborative, 250 First Avenue, Suite 300, Needham, MA 02494. Bio-IT World is free to qualified life science professionals. Periodicals postage paid at Boston, MA, and at additional post offices. The one-year subscription rate is \$199 in the U.S., \$240 in Canada, and \$320 in all other countries (payable in U.S. funds on a U.S. bank only).

POSTMASTER: Send change of address to Bio-IT World, P.O. Box 3414, Northbrook, IL 60065. Canadian Publications Agreement Number 41318023. CANADIAN POSTMASTER: Please return undeliverables to Station A, P.O. Box 12, Windsor, ON N9A 6J5.

Subscriptions: Address inquiries to Bio-IT World, P.O. Box 3414, Northbrook, IL 60065 (888) 835-7302 or e-mail biw@omedia.com.

Reprints: Copyright © 2009 by Bio-IT World All rights reserved. Reproduction of material printed in Bio-IT World is forbidden without written permission. For reprints and/or copyright permission, please contact the YGS group, 1808 Colonial Village Lane, Lancaster, PA; 717-399-1900 X 125 or via email to ashley.zander@theYGSgroup.com.

Bio-IT World®

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

EDITOR-IN-CHIEF

Kevin Davies (781) 972-1341
kevin_davies@bio-itworld.com

MANAGING EDITOR

Allison Proffitt (781) 972-1345
aprofitt@healthtech.com

ART DIRECTOR

Mark Gabrenya (781) 972-1349
mark_gabrenya@bio-itworld.com

VP BUSINESS DEVELOPMENT

Angela Parsons (781) 972-5467
aparsons@healthtech.com

VP SALES – WESTERN US, CANADA, EUROPE, PACIFIC RIM

Alan El Faye (213) 300-3886
alan_elfaye@bio-itworld.com

REGIONAL SALES MANAGER – NEW ENGLAND, NORTH EASTERN US, SOUTH EASTERN US, MIDWEST, INDIA

Kay O. Christopher (860) 693-2991
kchristopher@healthtech.com

SENIOR DIRECTOR OF MARKETING & OPERATIONS, PUBLICATIONS

Joan A. Chambers (781) 972-5446
jchambers@healthtech.com

PROJECT/MARKETING MANAGER

Lynn Cloonan (781) 972-1352
lcloonan@healthtech.com

ADVERTISING OPERATIONS COORDINATOR

Matt Gross (781) 972-5481
mgross@healthtech.com

PRODUCTION MANAGER

Tom Norton (781) 972-5440
tnorton@healthtech.com

Contributing Editors

Michael Goldman, Karen Hopkin, Deborah Janssen, John Russell, Salvatore Salamone, Deborah Borfitz Ann Neuer, Tracy Smith Schmidt

Advisory Board

Jeffrey Augen, Mark Boguski, Steve Dickman, Kenneth Getz, Jim Golden, Andrew Hopkins, Caroline Kovac, Mark Murcko, John Reynders, Bernard P. Wess Jr.

Cambridge Healthtech Institute

PRESIDENT
Phillips Kuhl

Contact Information

editor@healthtech.com

250 First Avenue, Suite 300
Needham, MA 02494



Successfully delivering IT and consulting services to world leaders in R&D, science and engineering



Innovative Informatics and Consultancy for R&D

- Greater efficiency and productivity in drug discovery
- Delivering the benefits of translational medicine
- Faster, more effective clinical trials

Tessella would like to congratulate Wyeth Research for winning the Bio-IT World Best Practice Award 2009

www.tessella.com
info@tessella.com

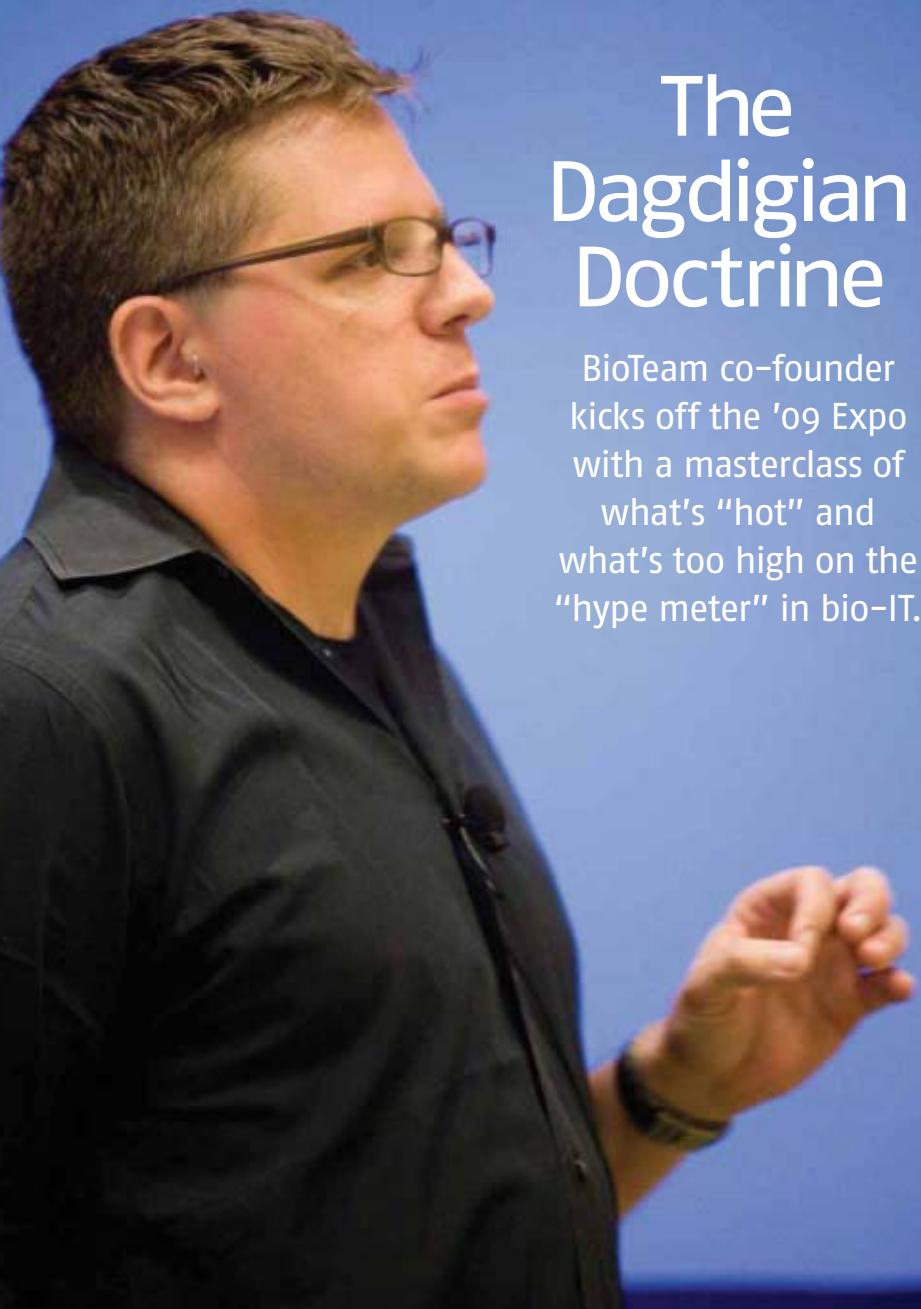


Up Front 2009 Bio•IT World Conference & Expo

SPECIAL REPORT

The Dagdigian Doctrine

BioTeam co-founder kicks off the '09 Expo with a masterclass of what's "hot" and what's too high on the "hype meter" in bio-IT.



MARK GABREYNA

BY KEVIN DAVIES

BOSTON — In the opening keynote of the 2009 Bio-IT World Conference & Expo, Chris Dagdigian delivered a candid assessment of the best, the worthwhile, and the most overhyped information technologies (IT) for life sciences.

Some formerly hyped technologies were now mainstream, including virtu-

alization and storage. Others were over hyped but still on balance worthwhile, such as green IT and utility computing. And in the future, Dagdigian saw major benefits accruing from trickle-down best practices and federated storage.

Dagdigian, a founding partner for the **BioTeam** IT consultancy, is a regular speaker at the annual Bio-IT World Expo. This year, he brought his trade-

mark "trends in the trenches" talk to the opening plenary session. Offering his customary disclaimers, he said he was "comfortable with deliverables, but not comfortable with being a talking head or a pundit." The audience begged to differ.

Already Mainstream

Dagdigian divided his talk into three sections: reviewing what he called "old news," discussing "currently exciting" technologies, and those that could be exciting in the future. But, he warned, there is "a ridiculous list of stuff on the 'hype meter.'

In the "already mainstream" category, Dagdigian tagged virtualization and the "bio-IT storage tsunami." Most of his recent work has been in helping "people getting buried by instruments." The community had been talking about the "data deluge" for the past four years. It is "still buried," said Dagdigian, "but the problem domain is understood. Our smallest customers aren't losing hope, and our biggest customers are staying ahead."

Last year, Dagdigian saw the first "100-Terabyte (TB) single namespace project." But, he cautioned, he had also witnessed for the first time a "10 TB catastrophic data loss" with consequent job losses. The loss, which occurred in a government lab, resulted primarily from double-disk failure in a RAID5 volume holding SAN F5 metadata. Dagdigian said he used to be a "huge fan of RAID5," but abandoned it last year. "The statistical probability [of failure] is too high—it's going to happen!" Everything is now RAID6, with maximum attention paid to monitoring and maintaining data integrity.

In a welcome development, data triage discussions are spreading beyond cost-sensitive industry organizations. "Everyone has come to the realization that data triage is a given," said Dagdigian. Displaying his favorite slide, Dagdigian showed a screen shot of an 82-TB folder on a Mac. Even more impressive was a 1-Petabyte (PB) output mounted on a Linux system output.

While storage is cheap (and getting cheaper), operational costs, staff, tape and backup costs are still fairly constant.

(CONTINUED ON PAGE 13)

Eric Schadt Leads Network News

There are no linear pathways, only networks says Sage co-founder.

BY KEVIN DAVIES

BOSTON—Eric Schadt, the Rosetta/Merck biomathematician in the process of setting up a new non-profit organization called Sage (see, “Merck Execs Tout Sage,” *Bio•IT World*, Mar 2009), gave a powerful testament in his keynote address to the need to focus on non-linear biological networks, rather than linear pathways, as the future of disease research and drug discovery.

Dressed in his standard white polo shirt and khaki shorts, Schadt explained that in 2007, genome-wide association studies (GWAS) were “all the rage,” rapidly producing more than 200 replicated loci for a host of complex diseases. But Schadt’s team took one of those candidate genes—*ERBB3* in type 1 diabetes (T1D)—identified in 2007, and asked: what was the functional support for these gene identifications?

“Is [the marker] a surrogate or is it the causal mutation? Does the variation activate or inactivate the gene? Can the gene be inferred from these studies? This can take years to figure out,” said Schadt, pointing to the example of Apolipoprotein E and Alzheimer’s disease.

To dissect these diseases, Schadt said, “we need to leverage all the biological data... Our brains appear to be wired for storytelling, not statistical uncertainty. The truth is, we have little idea on the underlying causes of disease.”

Pathways Out, Networks In.

Schadt argued that there is no such thing as a tidy linear pathway, only networks. That puts a premium on generating large-scale expression data, filtering GWAS data on expressed single-nucleotide polymorphisms (eSNPs) and causal networks to identify candidate genes and produce predictive network models. In this way, Schadt’s team identified a gene called *RPS26* as the T1D susceptibility gene



Forget single genes, for Schadt it's all about the pathways, tissues, and sub-networks.

first mapped in the 2007 GWAS. The expression of *RPS26* showed the greatest association of genes in the region with the variant, and KEGG analysis showed the major pathway enriched was indeed type 1 diabetes. Schadt praised Microsoft’s Amalga Life Sciences (see p. 58) as one of the significant efforts tackling this issue, and Gene Network Sciences (Schadt is an advisor) as a good example of integrating data and building models.

Gene expression patterns do a good job of refining clinical phenotypes, as Schadt showed himself several years ago in collaborating with the Netherlands Cancer Institute to develop gene signatures for breast cancer recurrence risk. “Can we go beyond correlation to find causal genes?” he asked. He expressed incredulity that, a few years ago, “drug companies made \$100-million bets based [solely] on gene expression correlations.”

Schadt’s major emphasis of late has been modeling causal relationships between and within different tissues. Ongoing work looking at gene expression data from mouse tissues is used to build networks reflecting interactions between

those tissues. In this way, Schadt’s team has found that highly interconnected sub-networks are connected *between* tissues. Cross-tissue networks are specific to those interactions, and would be invisible if one was focusing on any single tissue. “You’re missing those networks that are facilitating connection between tissues and disease processes,” said Schadt. “We want to get away from the single gene view of diseases, and view these sub-networks.”

Having identified DNA variants on chromosome 17, Schadt used RNA interference to target those sub-networks. “We punch genes in the liver network and see if that induces changes in the islet network,” he said. From there, his colleagues can design drug candidates targeting the network node. Ongoing work renders mice immune to becoming obese despite being on a high-fat diet. That drug is now moving into the clinic.

Schadt is transitioning out of Merck this summer to build Sage, where much of this data-intensive focus on biological networks will continue to grow. “It’s the network that’s sensing the variation and driving disease,” said Schadt. •

Up Front 2009 Bio•IT World Conference & Expo

Will the Gene Microscope Change the World?

Clifford Reid predicts a paradigm shift in genome sequencing.

BY ALLISON PROFFITT

BOSTON—The invention of the light microscope in 1590 advanced many scientific disciplines, but medicine, curiously wasn't one of them. Three hundred years later, a series of closely timed microscope advancements proved the tipping point and tuberculosis was all but eradicated in three years.

"We're building gene microscopes," said *Complete Genomics'* CEO Clifford Reid. "With this technology and the computer technology available to us, we're going to change the world."

In the final keynote, Reid attributed the genome sequencing explosion to the recent arrival of a series of disruptive technologies. "The slope of the line changed from 2x a year improvement [in sequencing costs], to 10x a year improvement. And the price of human genome sequencing began to collapse. Now if you extend that line... you'll realize that you'll be able

to get your genome sequenced in 2015 for one penny. This will not happen, I assure you."

Reid pointed to three areas of technology that have prompted the disruption: bio, nano, and info.

"In the bio area, molecular tools have exploded in the past five years; they have become extraordinary," he said, citing advancements in oligos, DNA amplification, fluorescent chemistry, and enzymology. "Coupled with that, we have this mas-

sive revolution in nano tools," including photolithography, robotics, and optical systems that are inexpensive, high quality, and innovative. Finally, "the revolution that we've all been living with... is the info technology revolution. There's been a fundamental informatic insight into DNA sequencing around paired end sequencing."

The last insight is one on which Complete Genomics has built their business. They start by circularizing a DNA fragment, adding read addresses at regular intervals around the circle, and then amplify the DNA into nanoballs (see, "Complete Genomics..." *Bio•IT World*, Nov 2008). "The extraordinary thing

tion by this time next year, it will take a half a day. Half a day to read a complete human genome worth of DNA."

Whereas chemistry costs used to drive the industry, Reid said that had changed. Today, the limiting factors are visualization and compute power.

Paradigm Shift

With these disruptive advancements, Reid sees a paradigm shift happening in the life sciences. "We can afford large scale human sequencing," he says. The old paradigm produced gigabytes of data that could be analyzed with a single computer. Now genomes consist of 60 terabytes of data and require high performance computing for assembly.

The new paradigm calls for new vendor and user models that CGI is spearheading. CGI will be offering information services, not a box. Sequencing should be scaled up to continuous production. Users, on the other hand, will outsource data generation so they can focus on data analysis and interpretation. In this way, CGI will be a wholesaler of complete human genomics to multiple users, and plans ten centers around the world with the goal of sequencing 1 million human genomes in the

next five years.

Reid's vision opens doors for large-scale studies in functional genomics, orphan disease research by academia, and cancer research. In the past, we were instrument-centric, Reid said. Today we're data-centric, and in the future we will be action-centric, finally realizing personalized diagnostics and personalized therapeutics.

The technology is now light years beyond the light microscope. •



Clifford Reid and Complete Genomics believe large-scale human genome sequencing is at hand thanks to disruptive technology.

about this amplification strategy is that it takes place in a test tube to amplify a genome's worth of sequence in about 20 minutes, and at today's enzyme prices costs about \$27."

From there, the DNA is dropped onto a slide to be read with fluorescent chemistry. Currently, Complete Genomics can read about 10% of a human genome sequence on a single slide in about a day and a half. "To read the next generation of slides that will be in commercial produc-



MARK GARBREY

The 2009 Bio-IT World Expo panel on The Future of Personal Genomics (l to r): Clifford Reid, Philip Reilly, Dietrich Stephan, Robert Green, John Halamka, Jorge Conde, and moderator, Kevin Davies.

Responsibility in Personal Genomics

Making genomics useful to the consumer is proving a challenge.

BY ALLISON PROFFITT

BOSTON — At the Bio-IT World Expo's second annual plenary roundtable on The Future of Personal Genomics, the discussion centered on the present and future delivery of genomic information to the patient. With personal genomics companies reaching out to consumers, new questions are raised about communicating results to individuals.

Guest panelists were Robert Green, [Boston University Schools of Medicine and Public Health](#); John Halamka, CIO, [Harvard Medical School](#); Dietrich Stephan, co-founder, [Navigenics](#); Jorge Conde, CEO, [Knome](#); Philip Reilly, [Third Rock Ventures](#); and Clifford Reid, CEO [Complete Genomics](#). The panel was chaired by *Bio-IT World* editor-in-chief, Kevin Davies.

"We want to be these folks' back offices," said Reid, of the consumer genomics

companies, "but they need to do analysis and counseling." And that is proving almost as tricky as the sequencing itself.

Since early 2008, Knome has offered whole-genome sequencing to wealthy clients. The company is learning to deliver results face to face, said Conde. "There's a lot we don't know about the genome yet," he admits. "How do we deliver this information in a responsible way, in a way that provides context and that can be digested by somebody or by individuals that aren't trained geneticists?"

Conde assembles a team of genetic counselors, scientists, and doctors, flying clients to Boston to deliver results over a full day. "We want people to be able understand their own information. We don't want to scare them. We don't want to mislead them; and we want to do it in a way that we're learning as well," he said.

John Halamka has a foot in both camps, as both a guarantor of medical information and a founding volunteer (patient number two) in George Church's Personal Genome Project. Halamka stressed that patients must understand probabilistic versus deterministic results. But this isn't something doctors can do

alone, he said. They need to be leveraging other experts.

Robert Green led the REVEAL trial, first empirical study of people's reactions to genomic data. The findings suggested that people who test positive for the high-risk Alzheimer's *APOE4* variant buy more long-term disability insurance, more vitamins and supplements, and are subject to what he calls "flat out pseudoscience." This data can be exploited, he warns.

However, Philip Reilly, a clinical geneticist, lamented the still woefully low numbers of genetic counselors in the U.S. Very few internists are competent to interpret and explain genetic data, he said, and what we know about linking genotype and phenotype is still "awful." Consumers need "meaningful, low cost ways to respond to genetic information," he challenged; otherwise, the data are of no use. Reilly advised against the word "predictive," suggesting "risk assessment" as a substitute.

Green, who describes himself as the world's oldest genetics resident, agreed. The public is consumed with low probability events, he said, and could overwhelm our current health systems. •

Up Front 2009 Bio•IT World Conference & Expo

Perfect Ten

Bio•IT World announces winners of the 2009 Best Practices Awards.

The Best Practices Awards Program recognizes academic and industry organizations for outstanding examples of innovation and collaboration in the deployment of technologies, and novel business strategies that will advance the enterprise.



Bio•IT World editors assembled a panel of expert judges, who reviewed a record 72 submissions from large pharmaceutical and biotechnology companies, academic institutions, and niche service providers. Those vigorous discussions produced eight winning entries and two discretionary awards—the Judges' Prize and the Editors' Choice Award.

"These annual awards spotlight some of the tools and strategies that are making a difference in the conduct of basic and clinical research, speeding the drug discovery pipeline, and changing the way we generate, share and analyze data and knowledge," said Kevin Davies, editor in chief of *Bio•IT World*. "We hope these outstanding examples of technology innovation and collaboration will drive future discoveries across basic, pharmaceutical, and clinical research."



The 2009 Bio•IT World Best Practices Awards winners:

Basic Research: NATIONAL CENTER FOR GENOME RESOURCES—NCGR's *Schizophrenia Genome Project*

Clinical Trials Management: GENENTECH—*The Clinical Trial Portal* (Nominated by ePharmaSolutions)

Clinical Trials Design: WYETH RESEARCH—*Design of adaptive clinical trials using Adaptive Design Explorer* (Nominated by Tessella)

Drug Discovery & Development: AMGEN and GENEDATA—*Amgen Lead Discovery Informatics*

IT & Informatics: The BROAD INSTITUTE of MIT and HARVARD—*CellProfiler*

Knowledge Management: VANDERBILT UNIVERSITY—*A computerized, clinically intelligent system to deliver clinical alerts to physicians and their patients improves care and lowers health care costs* (Nominated by ActiveHealth Management)

Knowledge Management, Pharma: MERCK & CO and ABBOTT LABORATORIES—*Biological Knowledge Management: Registration, Association, and Sharing* (Nominated by Accelrys)

Translational and Personalized Medicine: GLAXOSMITHKLINE—*SAfety Works: Leveraging Observational data to explore the effects of medicines* (Nominated by ProSanos Corporation)

Judges' Prize: CHILDREN'S HOSPITAL OF PHILADELPHIA—*Pediatric Knowledgebase*

Editor's Choice Award: ASTRAZENECA—*Safety Intelligence Program* (Nominated by BioWisdom)

Complete coverage of the 2009 award winners, including interviews with all of the winners, details of the entries, and a report of the awards ceremony and dinner talk by Michael Cariaso (BioTeam) will appear in the next issue of *Bio•IT World*.

Dagdigian

(CONTINUED FROM PAGE 8)

Users, enamored with the plummeting price of TB storage devices from Costco, do not understand enterprise IT and backup requirements. “IT organizations need to set expectations, because the electronics market is skewing expectations,” said Dagdigian. Many next-generation sequencing grants simply didn’t budget for storage, let alone a 100+ TB storage system. Meanwhile, Dagdigian noted, the [Broad Institute](#) already has more than 1 PB of storage.

“Unlimited data storage is over,” noted Dagdigian. It’s simply not possible to back up all data, keep it safe, secure, and so on. “Sometimes,” he said, “it’s better to go back to the -40 F freezer and repeat the experiment.”

In short, storage is no longer a major bottleneck—rather, that falls to chemistry, reagents, and human factors. Customers are starting to trust instrument vendor software more. “The problems are not as scary as they once seemed,” he said. Dagdigian also noted that storage devices are running more 3rd party software, such as [Ocarina](#) Reader software on [Isilon](#), and Ocarina Optimizer on [BlueArc](#).

Virtualization offered the “lowest hanging fruit,” said Dagdigian. The tipping point, Dagdigian said, was the live migration of a VMS (virtual memory system) without requiring a proprietary file system underneath. In 2009, he helped build and design a virtual collocation facility for an academic west coast campus, which was experiencing limits imposed by electrical power and air conditioning. “400 servers are currently virtualized on a lightweight simple platform,” said Dagdigian. Large numbers of physical servers have been shut down, realizing significant savings from de-duplication, compression and thin provisioning, not to mention electricity. Moreover, scientists now have full administration control.

Coming soon, said Dagdigian, “virtualized cluster head nodes.” Not coming soon: grids and clusters distributing entire VMS for task execution. It isn’t practical, argued Dagdigian but rather a case of “marketing winning out over practical stuff.”

Bourne Free Access Nets Franklin Award

Philip Bourne ([UC San Diego](#)), winner of the 2009 Benjamin Franklin Award presented by the Bioinformatics Organization, championed “open science” in his laureate lecture, while admitting there had to be a business model around it. “Someone has to pay for it,” said Bourne.

Academic research was breaking out of the traditional cycle of grants/posters/papers/journal articles, thanks to database curation, blogs, wikis, even multimedia (see p. 5). The distinction between knowledge and data is merging, Bourne said, with journals publishing more supplementary information, and databases increasingly curated. Bourne has even helped form a society for biocuration.

Bourne pointed to the value of semantic enrichment to papers, with web service calls to various databases for associated data. Another goal revolves around disambiguation. “I want the Bourne Identity!” he said—the equivalent of a DOI (digital object identifier) for authors. Initiatives such as Open ID and the NCBI’s MyBibliography are helping.



MARK GABREY

Green IT

In the category of “hyped beyond all reasonable measure—but still worth pursuing,” Dagdigian said green IT could deliver real electrical savings. “Use green IT for political cover,” he urged the audience. A deployment of a Nexan SATABeast had led to a 30% reduction in power draw with no impact on cluster throughput. One of his best 2009 moments came in deploying a Linux HPC cluster for a west coast organization. The system interface talks to Platform LFS, powering nodes up and down and sending automatic email alerts to management such as: “Hello, I’ve saved \$80K in facility costs this year.”

Utility (or cloud) computing is “not rocket science, but fast becoming mainstream,” said Dagdigian. “[Amazon](#) web services [EC2] is the cloud,” he said. “It’s simple, practical, and understandable,” and enjoyed a multi-year head start on the competition. The rollout of features are “amazing,” such as Hadoop and applications for short-read sequence mapping. “I drank the EC2 Kool-Aid: I saw it, I used it, I solved real-world problems,”

said Dagdigian.

The biggest problem is ingesting data into the cloud. “There is no easy solution,” he said. How does one push 1TB/day into Amazon? “Have patience,” said Dagdigian, expressing “100% confidence” that Amazon is working on the problem. “If we can solve data ingestion problem, I see a lot of scientific data taking a one-way trip into the cloud. Data would rarely, if ever, move back. If I take it back, it’s going to be really obnoxious—big data pipes or people driving minivans of USB drives.”

Worth Watching

In his “worth watching” category, Dagdigian cited federated storage and the trickle-down of best practices from Amazon, Google, and others. Amazon, Google, and Microsoft had all been computing at such a scale that it was too much of a trade secret, he complained.

But there are signs, such as a recently released video of the Google data center circa 2004, that their best practices will eventually trickle out, benefiting the entire community. •

Up Front 2009 Bio•IT World Conference & Expo

Best of Show Winners

The annual contest to recognize the most innovative and important new products on the exhibit floor.

Informatics Tools & Data: CLC bio

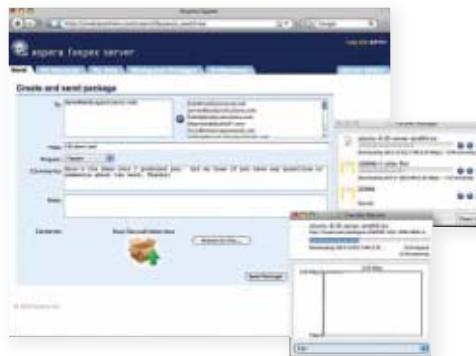
CLC Genomics Server is an advanced and powerful bio-informatics solution including a powerful and modern three-tier server architecture, flexible options of executing centralized services, easy integration with other applications and services, powerful database communication and data integration, and secure access control framework and central action logging. The solution provides a user-friendly and fast backbone for any enterprise platform within bioinformatics and next-gen sequencing data analysis.

Product: CLC Genomics Server

Company: CLC bio

Available: Now

For more information: www.clcbio.com



IT Infrastructure: Aspera

The **Aspera** Enterprise Server serves up files, directories, and large data sets with unrivaled performance. Aspera Enterprise Server is built on the company's fasp transport architecture, an innovative, ultra-reliable protocol designed to meet the speed, control and security requirements of business-critical file transfer over any IP network (LAN, WAN, satellite and wireless). fasp transport delivers predictable, guaranteed delivery times regardless of network conditions. Enterprise Server also features comprehensive user authentication, unlimited numbers of users, comprehensive monitoring and reporting, and more.

Product: Aspera Enterprise Server

Company: Aspera

Available: Now

For more information: www.asperasoft.com

Entrants/Finalists

* **Accelrys' Pipeline Pilot** is a data integration, transformation, analysis, and reporting platform. The most recent version integrates SharePoint and Pipeline Pilot by allowing Pipeline Pilot protocols to be deployed to, and executed from, SharePoint pages. www.accelrys.com

Bioalma's novoseek is a powerful alternative to tradition information retrieval systems. It indexes biomedical literature with text mining technology that enables users to identify key biomedical terms and take into consideration contextual information (see, p. 29). www.novoseek.com

* Best of Show Finalist

* **BioWisdom's Safety Intelligence Program** is a combination of innovative web-based search technology with key quality knowledge in a unique "fact-based" format known as assertional metadata (AMD). It represents the largest collection of drug molecular mechanisms. www.biowisdom.com

BlueArc's 6.x release debuted several new file system features unique to the storage market including the combination of the Data Migrator, Dynamic Caching, and Enhanced Cross-Volume Links products. These three features form the backbone of the ILM (Information Life-cycle Management) solution. www.bluearc.com

CycleComputing's CycleCloud quickly de-

livers scalable, enterprise-class high performance compute clusters in the Cloud. CycleCloud delivers secured cluster scheduling and management features to cloud infrastructures like Amazon's EC2. www.cyclecomputing.com

DNASTAR's ArrayStar v3.0 is an expression analysis software package that permits users to work with either microarray or sequence data for expression characterization projects. 3.0 adds analysis and visualization features to the Gene Ontology section. www.dnastar.com

Good Products' g-docs is an electronic document management solution designed to help organizations in a compliant environment cope with increas-

Knowledge Management: Rescentris

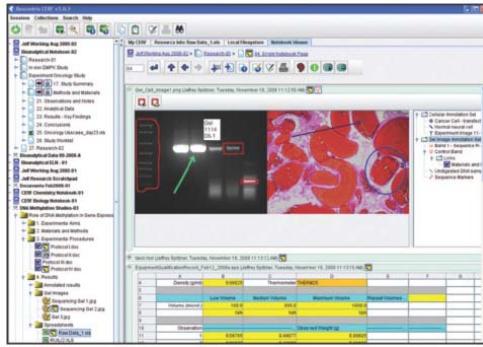
Rescentris' Collaborative Electronic Research Framework, or CERF, is a 21CFR11-compliant, cross platform electronic lab notebook solution that enables better management of lab records and scientific content. CERF 3.5 offers new options for data hosting and archiving and new project management features that allow researchers to maximize productivity. CERF 3.5 uses flexible, highly scalable architectures where research data is recorded on central servers accessible from any location and includes a new Software as a Service option and new secure digital archiving solutions.

Product: CERF v3.5

Company: Rescentris

Available: Now

For more information: www.rescentris.com



ingly complex regulations. G-docs can track versioning, manage approvals, and monitor and record changes in an audit log. www.goodproductsLtd.com

The HP StorageWorks 9100 Extreme Data Storage System provides vast storage capacity and simplified, integrated management with a small footprint. The base storage configuration provides three high-availability storage blocks and 246 TB of capacity. www.hp.com

* **IMC's SAFR NS** generates real-time efficient tracking of safety reports, enabling reporting and compliance. SAFR NS uses AJAX, integrates with Microsoft Exchange mail servers, and uses load balanced application server instances. www.imc.com

Indigo BioSystems' Rubicon achieves data integration in large-scale, distributed research and development. The system has four components based on open source standards: a raw data repository, information integration, web-based workflow framework; and a hosted data interchange service. www.indigobio.com

Infinite Peripherals' AssetRelay manages assets by Bluetooth or USB barcode scanning signature capture, or GPS. The solution is offered as software or software/hardware options with a graphic BlackBerry user interface. www.ipcprint.com

Infosys' Cell Line Registration and Analysis system is a prototype for biological reg-

istration system. Features include: cell line data schema, functional workflows, transactional workflows, search capabilities, and Tibco-Spotfire service for visual inspection. www.infosys.com

Ingenuity has added contextual data analysis to **Ingenuity Pathway Analysis**. Contextual data analysis is comprise of a series of powerful biological filters and other tools that make it easier to quickly understand experimental results and fine-tune scientific hypotheses. www.ingenuity.com

* **IO Informatics' Sentient Suite** integrates data, applications, databases, and instruments into one secure and compliant interoperable environment. Users can access, query, and use data from multiple



Clinical Trials & Research: MaxisIT

CT Renaissance is offered as a Software as a Service, enterprise-class, completely integrated and automated software suite for clinical trials offering the most efficient web-based software solutions that involve no IT footprint at the point of use and comply with all regulatory requirements. With an XML backbone, CT Renaissance offers customizable functionalities to manage world-wide clinical trials including collaborative writing and reviewing, statistical computing environment, trial management, centralized data repository, electronic submission, and business process engine.

Product: CT Renaissance

Company: MaxisIT

Available: Now

For more information: www.maxisit.com

Up Front 2009 Bio•IT World Conference & Expo

Best of Show Entrants/Finalists

sources regardless of hardware or IT infrastructure. www.io-informatics.com

* **Isilon IQ with OneFS 5.0** is the first and only storage system to scale to greater than 3.4 Petabytes of capacity and up to 30 Gigabytes per second of aggregate performance from a single file system and single volume achieving more than 200X the scalability and 60X the performance of traditional SAN and NAS storage systems.

www.isilon.com

NextBio provides a life sciences search engine and discovery platform, which enables researchers and clinicians to better access, understand, and share data. NextBio indexes and correlates highly complex experiments, literature, and clinical trials to enable discovery. www.nextbio.com

* **Ocarina Networks' ECOSystem** includes the Optimizer, which compresses data and files, and the Reader, software that integrates with file servers to decompress the files. The system uses a “context-aware” approach that optimizes compression based on patterns within the data.

www.ocarinanetworks.com

ParaScale is a software-only solution that can be downloaded from the web and applied to any Linux platform. ParaScale enables hundreds of commodity servers to be clustered together to act as a file repository with massive capacity and parallel throughput.

www.parascale.com

PBS GridWorks' Personal PBS uses the same technology as PBS Professional to make real-time, on-the-fly adjustments based on processor and memory availability. It allows users to work on their desktops while submitted jobs run in the background.

www.openpbs.org

* **Phase Forward's Clarix Interactive Response Technology** is a next generation

* Best of Show Finalist



MARK GABREY

interactive voice/Web response system that operates on a single, unified platform for telephone and Web. The platform provides system robustness and improved risk management to support global clinical trials.

www.phaseforward.com

PleaseReview makes the process of collaborative review and authoring less painful and more efficient within a department, across an enterprise, inside or outside of firewalls.

www.pleasetech.com

Plectix' Celluciate is a new computational approach to cell signaling that provides scientists with a collaborative, web-based workspace for simulating cell signaling networks and exploring their causal and dynamic properties.

www.lectix.com

QUOSA Literature Workflow Solution provides a complete literature access environment for the enterprise including archiving and metadata, supporting both those managing and vetting literate resources on behalf of others and end users.

www.quosa.com

Selvita's Protein Modeling Platform is a web-based modeling tool, built on the idea that protein modeling can be more user-friendly and practical. The platform features superior protein structure prediction at all the difficulty

levels based on CASP top competitors technology.

www.selvita.com

* **Sophic's Cancer Biomarker Knowledge Network** has converged the NCI Cancer Gene Index, release three of the Biomax BioXM Knowledge Management environment, and the Cancer Biomarker Knowledge Network. The database is NCI-funded and open source.

www.sophicalliance.com

TriLink Biotechnologies has released **CleanAmp dNTPs**, the newest addition to TriLink's line of PCR-enhancing products. CleanAmp products offer a complete chemical solution to Hot Start activation through primers and dNTPs to control mis-priming and primer dimmer formation.

www.trilinkbiotech.com

Valiance's TRUcompare is the only commercially available automated testing tool for data and content migration. The testing process uses configurable source to target data transformation to enable 100% testing, reducing time and cost.

www.valiancepartners.com

Xyntek's ITx-cGene Assay Data Analyzer & Regression Curve Generator is highly customizable Java and Oracle-based application for regression-curve fitting and analysis on dose-response data maintained in a highly scalable database.

www.xyntekinc.com

Up Front News

The Vagaries of Genome Variation: Do You Copy?

Comprehending Copy Number Variation makes an impressive debut.

BY KEVIN DAVIES

SAN DIEGO—The inaugural *Comprehending Copy Number Variation* conference* featured many of the true pioneers in the field, and left attendees impressed not only with the sheer number and breadth of copy number variants (CNVs) in the human genome, but also emerging evidence for their role in the etiology of many complex diseases, including autism, epilepsy, and schizophrenia.

Re-discovered by two groups in 2004, CNVs are stretches of DNA of greatly differing size that may be duplicated, deleted or inverted in different genomes. The amount of genetic variation caused by CNVs is 7-10 times more than the 0.1% attributed to single nucleotide polymorphisms (SNPs). CNV study has exploded thanks to comprehensive CNV

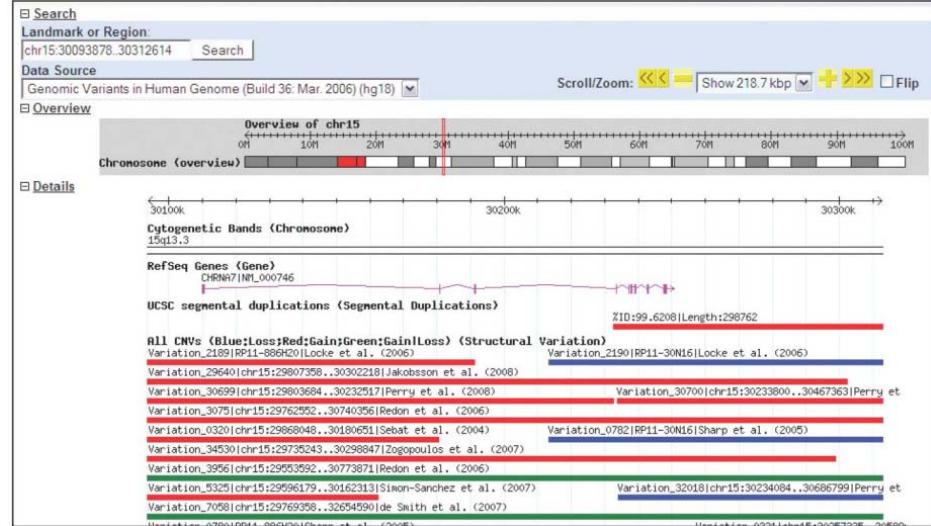
arrays from the likes of Agilent, Nimblegen, Illumina, and Affymetrix, although **Jan Korbel (EMBL)** noted that no single platform detects all structural variants. His group has found widespread CNVs in the human olfactory receptor genes.

Stephen Scherer (Hospital for Sick Children, Toronto) reviewed the current state-of-play, as chronicled at the Database of Genetic Variants (projects.tcag.ca/variation), which currently catalogues more than 21,000 CNVs at more than 6500 loci. “We’re leaving Kansas,” Scherer said, stressing



Evan Eichler

*CHI's Comprehending Copy Number Variation. San Diego March 17-18, 2009.



Online repositories such as the Database of Genetic Variants catalog the thousands of CNVs discovered so far, including many linked to complex diseases.

that at least one third of human genes exhibit some sort of sequence variation. Some of the best characterized examples include *SHANK3* on chromosome 22 and *ANKRD11* on chromosome 16, implicated in several cases of autism.

Evan Eichler (HHMI/University of Washington) showed examples of CNVs with clinical relevance. A common inversion of a region on chromosome 17 in northern European populations acts as a pre-mutation state associated with mental retardation. Another 1.5-megabase deletion on chromosome 15 occurs in 0.2–0.3% cases of autism, mental retardation, and schizophrenia, and up to 1% epilepsy.

Eichler’s group is identifying CNVs in individual genomes by simply looking at the genome sequencing depth of coverage among the millions of sequence reads. First, they developed a new mapping

algorithm that, unlike programs such as MAQ, could map reads into repetitive genomic regions. He pointed to a 70,000 stretch of DNA that exists in two copies in Craig Venter’s genome but three in James Watson’s. “One can only imagine what that type of sequence might do,” Eichler joked. The group can now accurately assess how many copies of a gene there are in a region in a more quantitative fashion than standard array comparative genome hybridization methods. “I’m very excited by this. I think this is the first time that we can look at duplicated regions of the genome and accurately assess copy number based on depth of coverage.”

Harvard Medical School’s Charles Lee presented remarkable variation in the amylase genes among people of different ethnic backgrounds. Lee’s group found that populations with high starch intake (European, Japanese) had more copies of the amylase 1 gene than other populations. Japanese subjects even had copies of the gene on a second chromosome.

Lee said there is a need for a revised human reference genome. “No-one will really admit that the human genome is not really completed,” he said. •

Mark Boguski's Search for Resounding Health

An online resource for intelligent disease management.

BY KEVIN DAVIES

Sensing a need to provide accurate health information, Mark Boguski is launching a new direct-to-consumer Internet search firm called **Resounding Health**.

"Resounding Health is much more than just another search engine," says Boguski. "It's a total vertical search environment for web-based medical and health information" that will evolve. Boguski likens it to the developing human brain, which prunes non-useful connections over time. "We generate hundreds of thousands of interconnections among medical concepts and topics and, over time, the poor-quality ones are pruned away through the results of users' searches and their ratings of sites."

As a physician, Boguski realized early on that he could potentially help more patients by making scientific discoveries than he could see in a lifetime of medical practice. For nearly two decades, Boguski has played important roles in the bioinformatics and bio-IT arenas, helping to establish GenBank and Rosetta Inpharmatics in the 1990s, before directing the Allen Institute for Brain Science (see "Keynoters Focus on Challenges," *Bio•IT World*, May 2004) and a two-year stint in charge of functional genomics at the Novartis Institute for Biomedical Research.

As he was leaving Novartis, he felt there was also an unrecognized medical need "to provide better health information not only to consumers but also to doctors who are rapidly falling behind in their knowledge of new discoveries and technologies that could enhance patient care."

Going Deep

The company was originally going to be called Vertical Health, emphasizing the search capability in the deep health space

compared with the horizontal searches of Google and Yahoo! When his lawyers discovered the name was taken, Boguski needed an alternative. During a concert by the Chicago Symphony Orchestra, he saw an ad in the program for the symphony's record label—*CSO Resound*. And so Resounding Health was born.

Boguski and Littleford have devised methods to compare the results of searches in a canonical way and use that information to guide other users' searches. "All of the searches are anonymous but we do provide social networking tools for those users who wish to interact with others who share their interests or conditions," says Boguski.

Boguski's goal is to create "value-adding process networks" to provide "the most cost-effective management of chronic diseases in the near future." When the Resounding Health "brain" achieves a critical mass of validated and important connections, "we will have in effect crowdsourced a revolutionary new, e-patient-friendly textbook of personalized medicine and customized care."

Resounding Health has been bootstrapped by angel investments so far. Boguski launched the Web site (www.resoundinghealth.com) in conjunction with his talk at the 2009 Bio-IT World Expo. He hopes to attract advertising as traffic builds, and explore other B2B ideas.

Boguski says Resounding Health might be "the most important thing I've ever done in my career in terms of being as a physician helping patients." While it won't provide the gratification of those 1-to-1 patient encounters, he hopes the site will lead to "more efficient and productive clinical encounters and other interactions between patients and health care systems." •

More Online

Resounding Health Webcast:
[www.bio-itworld.com/LSW/
 Mark_Boguski](http://www.bio-itworld.com/LSW/Mark_Boguski)



ILLUMINA ON A ROLL

Illumina Genome Analyzers have been widely distributed recently. In the past few months, the **Beijing Genome Institute** added 12 GAs to its fleet; the **Broad Institute** bought 20 additional instruments; and the **Genome Center at Washington University School of Medicine** in St. Louis acquired 21 genome analyzers.

HPC PARTNERSHIP

Panasas and **Dell** have announced a partnership to deliver an integrated, high-performance, clustered computing solution for life sciences. The solution is an optimized combination of Dell PowerEdge M610 blade servers based on the latest Intel Xeon processor 5500 series architecture, and high-performance Panasas ActiveStor storage.

CLOUD ANALYSIS

Applied Biosystems and **Geospiza** announced the launch of a next-generation genetic analysis solution on the **Amazon** cloud. The solution is the life-science industry's first jointly offered genomic analysis cloud-computing solution, which is an approach to computing in which dynamically scalable and often virtual resources are provided as a service over the Internet. This offering leverages the web-scale infrastructure from Amazon Web Services.

NAVIGATING CHANGES

Navigenics' management has been busy lately. Jonathan Lord has replaced Mari Baker as CEO; co-founder David Agus now heads **University of Southern California's** new Center for Applied Molecular Medicine; and fellow co-founder Dietrich Stephan hopes to build a \$245-million, 300,000-square-foot research institute, the **Ignite Institute for Individualized Health**, in Northern Virginia.

Inside the Box



Antibody Docking on the Amazon Cloud

BY ADAM KRAUT

It was 18 months ago in this column that my BioTeam colleague Mike Cariaso proclaimed, "Buying CPUs by the hour is back" (see, "Sunny Skies for Compute Cloud," *Bio•IT World*, Nov 2007), in reference to Amazon's Elastic Compute Cloud (EC2). Back then, we were perhaps a bit ahead of the hype vs. performance curve of cloud computing. A few forward-thinking companies were finding ways to scale out web services, but there was little EC2 activity in the life sciences.

However, in the past two years, utility computing has begun to make an impact on real world problems (and budgets) in many industries. For life scientists starved for computing power, the flexibility of the pay-as-you-go access model is compelling. The Amazon EC2 process makes the grant process used by national supercomputing centers look arcane and downright stifling. Innovative research requires dynamic access to a large pool of CPUs and storage.

A great place to clear the air about this emerging technology is computational drug design. IT and infrastructure decisions made early in the discovery process can have a profound impact on the momentum and direction of drug development. For protein engineers and informaticians at Pfizer's Biotherapeutics and Bioinnovation Center (BBC; see "Programming at Pfizer's BBC" *Bio•IT World*, Jan 2009), the challenging task of antibody docking presents enormous computational roadblocks.

Respectable models of a protein's three-dimensional structure can usually be generated on a single workstation in a matter of hours. But they require refinement at atomic resolution to validate whether the newly modeled antibodies will bind their target epitopes. One of the most successful frameworks for studying protein structures at this scale is Rosetta++, developed by David Baker at the University of Washington (see, "Improving Structure Prediction," *Bio•IT World*, Nov 2007). Baker describes Rosetta as "a unified kinematic and energetic framework" that allows "a wide-range of molecular modeling problems ... to be readily investigated." Refinement of antibody docking involves small local perturbations around the binding site followed by evaluation with Rosetta's energy function. It's

an iterative process that requires a massive amount of computing based on a small amount of input data. The mix of computational complexity with a pleasantly parallel nature makes the task suitable for both high-end supercomputers and Internet-scale grids.

Rosetta So Much Better

When Giles Day and the Pfizer BBC informatics team designed its antibody-modeling pipeline using Rosetta, it soon realized it had a serious momentum killer. Each antibody model took 2–3 months using the 200-node cluster. With dozens of new antibodies to model, the project essentially grid locked until the team could find enough compute capacity to do the sampling. Plus, the pipeline was used unpredictably as it hinged on results in other departments. What was needed was a scale-out architecture to support "surge capacity" in docking calculations.

Traditional options were limited to expanding in-house resources by adding more nodes to the cluster or reducing the sampling. A doubled CPU capacity could potentially halve a two-month calculation, but would entail acquisition, deployment, and operational costs. Consequently, Day contracted the BioTeam to provide a cloud-based solution.

The result was a scalable architecture custom fit to their workloads and built entirely on Amazon Web Services (AWS). The AWS team is years ahead of the competition. AWS is

**Innovative
research requires
dynamic access to
a large pool of
CPUs and storage.**

unveiling new features and API improvements almost monthly. The AWS stack is fast becoming our first choice for cost-effective virtual infrastructure and high-performance computing on-demand.

The new docking architecture at Pfizer employs nearly the entire suite of services

offered by Amazon. A huge array of Rosetta workers can be spun up on EC2 by a single protein engineer and managed through a web browser. As Chris Dagdigian pointed out in his Expo keynote (see p. 8), the cloud isn't rocket science. The Simple Storage Service (S3) stores inputs and outputs, SimpleDB tracks job meta-data, and the Simple Queue Service glues it all together with message passing. What Amazon did right in 2007 was elastic compute and storage. What they do even better in 2009 is to provide developers everywhere with a complete stack for building highly efficient and scalable systems without a single visit to a machine room. The workloads at Pfizer that previously took months are now done overnight and the research staff can focus on results without pushing their projects on the back shelf.

Adam Kraut is a Scientific Consultant at BioTeam. He can be reached at kraut@bioteam.net

Will Consumers Sustain Direct-to-Consumer Genomics?

BY KATHIE WRICK

Welcome to the world of consumer genomics, where there are different rules for building successful businesses than in medical diagnostics. A host of companies are marketing or selling genetic tests directly to consumers. As expected, some companies are far more evidence-based than others in their test and product offerings.

Personal genomics companies first launched in November 2007, including 23andMe, deCODEme, and Navigenics, and have brought the latest in gene chip technologies to the marketplace. Knome will sequence your entire genome for \$99,500. But genetic testing has been marketed to consumers in various ways for a long time. Most of the earlier firms built on past capabilities in doing forensic genetics and, as Internet retailing started to grow, began offering paternity and family relationship tests online to consumers.

The science and business have evolved rapidly in the past decade. Certainly, medical genetics will not completely evolve to a consumer business. But thanks to the confluence of two transformational technologies—the Internet, and the delineation and measurement of the human genome—procurement of some genetic tests has migrated from the scientist- or health professional-controlled domains to the world of cyberspace.

Entrepreneurs have many motivations for selling genetic tests directly to consumers. They have watched consumer disengagement with the U.S. health care system and its institutions grow exponentially for two decades, and consumers are now more proactive than ever in making their own health care decisions. Some firms believe that genomic medicine is coming anyway, and consumers armed with their genetic information

Further Reading:

Direct-To-Consumer Genetic Testing: Business Prospects in the United States, by Kathie Wrick, PhD, covers the business of genetic tests marketed and sold directly to consumers. Published by Insight Pharma Reports, March 2009. <http://www.insightpharmareports.com/genetic+testing.aspx>

can help drive it even faster. Others feel that everyone should be aware of their own genetic information, regardless of what the medical profession thinks. The ability to make expensive purchases securely online became the bridge that made consumer genetics happen, as prices for these tests or services range from hundreds to many thousands of dollars. Though medical genetic testing will grow in its own right, and much of this work and associated business revenues will stay within the traditional clinical laboratory testing services market, the Internet has allowed a mini “distribution revolution” for genetic tests.

The Biggest Variable

There is a big consideration for the technologically savvy genetic testing companies who have elected to sell their services directly to consumers—the consumers themselves. The 2007 startups have the best scientists and latest technologies. Yet it is not clear that personal genomics companies have applied the same level of rigor and resources to understanding their consumer marketplace as they have to the technology developments which helped make genetic testing for consumers affordable. Had these startups solicited funds from investors specializing in consumer products businesses, they might have been told to come back when they had appropriately sized their market and could back up an estimate of sales revenues based on sound consumer research. Consumer goods and services are different businesses altogether than medical diagnostics, pharmaceuticals, or health care. The compelling genetic technology advances applied to direct-to-consumer testing services would likely take a back seat to well-done consumer research that reveals how many consumers of a certain psychographic profile show a strong intent to purchase.

It's not clear that the current investors in personal genomics companies, who know technology businesses very well, are asking the right questions about what is important in a consumer products and services business. Even the publicly available surveys on consumer attitudes about genetics and genetic testing suggest strongly that consumers may not be the best targets for marketing and selling these tests. That's not to say there won't be a good number of consumers who will buy them. Rather, it just might mean that the vast majority of consumers, whose numbers are needed to sustain and grow a business long term, may not be likely to buy.

The secret could lie in identifying the consumer segment that is highly motivated to buy, learn what drives them, and design product offerings according to exactly what they are looking for—better than the competitors. The genetic test may not ultimately be the end product, but the vehicle for businesses to help consumers plan their lives after learning the results. Even then, without good consumer research, no one knows what the right product is or how big (or small) that product's business might be.

Kathie Wrick can be reached at kwrick@comcast.net.

Trends in Global Economic Development

The biotech and pharma industries, along with much of the world, are undergoing a season of reorganization and change. But government investments have shown that biotech is still one of the most sought after industries. Investments in biotech development continue around the United States and the world.

■ The Alexandria Center for Science and Technology at the East River Science Park in **New York City** will be a state-of-the-art research and development campus for biotechnology and pharma companies looking to locate in Manhattan. Despite the economic climate, venture capital funding for technology increased 10% (\$1.9 billion from last year) in the New York City metro area. 308 companies received VC funds.

■ **Delaware** recently won a five-year, \$17.4-million grant from the National Center for Research Resources (NCRR) at the National Institutes of Health (NIH). Six institutions partnered to win the grant: Christiana Care Health System, Delaware State University, Delaware Technical & Community College, Nemours/A. I. duPont Hospital for Children, the University of Delaware, and Wesley College.

■ The **San Jose** BioCenter has been named the world's top technology incubator by the National Business Incubator Association (NBIA) for the second year in a row. In three years, resident companies at the SJBC have raised more than \$700 million in VC funding from the country's top investment firms like Kleiner Perkins and Interwest, and top corporate investors like Sony and Samsung, all while attracting companies from across the globe to grow their business in San Jose.

■ The **Arizona** Biomedical Research Commission (ABRC) and SAM Solutions have entered into an agreement to develop the



state's first virtual tissue bank that will serve Arizona hospitals and researchers. This virtual tissue bank will provide a single, consolidated view of the tissue samples stored in repositories in a number of Arizona hospitals and research facilities. With this virtual tissue bank, researchers will be able to easily browse and query the virtual collections and determine the tissue samples that might be suitable to their research needs.

■ Baxter International announced the expansion of its **Singapore** manufacturing facility, with the addition of a biopharmaceutical facility. This facility will carry out bulk production for ADVATE, the recombinant factor VIII (rFVIII) therapy that is free of blood-based additives, to meet the growing needs of people living with hemophilia A. This investment will create about 230 new jobs at the Singapore facility.

■ The province of **Ontario** in Canada is investing \$4.9 million through its Biopharmaceutical Investment Program, part of the government's Next Generation of Jobs Fund. Purdue Pharma Canada, known for pione-

ring research on persistent pain and associated treatments, will be hiring 53 people to work in research, development and advanced manufacturing with the expansion of its operations in Pickering, Ontario. The company intends to break ground on the \$32 million expansion in January, 2009. The investment will double the company's R&D capacity in Ontario.

■ Site Selection has named **Missouri** the North American leader in the development of private-sector life sciences facilities. Missouri was recognized for 9 life science projects last year alone, was tied at the top with California and Pennsylvania. With more than 1,000 life science businesses located in Missouri and over 32,000 jobs in the field, the state has a strong base of support.

Putting the Alpha in Alabama

Jim Hudson preaches educational outreach and economic development at the Hudson Alpha Institute.

BY KEVIN DAVIES

While the spotlight on economic development in biotechnology is firmly on Georgia this May, as Atlanta plays host to BIO, there are exciting happenings just across the border in northern Alabama. According to various magazine surveys, Huntsville, Alabama, is one of the more desirable places to live and find work in these economic tough times. That goes for life scientists too, thanks in no small part to the efforts of Jim Hudson.

For a decade or more, Hudson ran Research Genetics, a supremely successful molecular biology service company in the 1990s that helped spur a golden age of gene discovery, which was eventually acquired by Invitrogen in 2000. Hudson and a close friend, whom he prefers not to name, decided to create "an institute that would embody that spirit of collaboration." The governor of Alabama, Bob Riley put up \$50 million to match the co-founders investment, and the Hudson Alpha Institute was born—the cornerstone of the 150-acre Cummings Research Park Biotech Campus.

The research operation is fairly modest for the time being, with five principal investigators led by former Stanford University genome center director Rick Myers, but Hudson expects the number to double or triple in the coming years.

What makes the Hudson Alpha experiment especially interesting is that it is also a catalyst for economic development and educational outreach. Since the end of 2007, the institute has been a hub for biotechnology entrepreneurship, providing a home for at least a dozen start-up companies. "We have a number of biotech companies in the same building," Hudson explains. "They all eat at the same place, go to joint seminars, have brain-storming sessions, etc."

"Boosting economic development—that's the mission," says Hudson, at least when it comes to the genomics part of life sciences in Alabama. "Genomics is a

paradigm that's uncomfortable to a lot of people, because it involves collaboration." Young investigators are often encouraged to prove they can do independent research, he says, but, "We look for people that understand the value of collaboration. We want people that are entrepreneurial in nature, that want to see their ideas wind up in the clinic, and potentially want to benefit financially from that as well." Myers, he says, "is not especially entrepreneurial, but he's not opposed to the idea!"

Many of the start-ups sharing space with the institute got their start when Invitrogen closed the Research Genetics site, prompting many employees to launch their own companies. One of those companies, Open Biosystems, picked up where Research Genetics left off and is now part of Thermo Scientific.

Another former Research Genetics spin-out is Applied Genomics (AGI), which



The Hudson Alpha Institute is the cornerstone of the 150-acre Cummings Research Park Biotech Campus.

was based on collaboration with Stanford microarray pioneer Pat Brown. AGI designed hundreds of antibodies to complement Brown's gene expression studies. MammoStrat is a diagnostic test that uses five monoclonal antibody biomarkers and a diagnostic algorithm to stratify breast cancer metastases, similar to Genomic Health's gene-based test. Expression Genetics (EGN) is enrolling patients for a gene therapy trial in ovarian cancer, based on a novel plasmid transfection reagent.

The third leg of the Hudson Alpha Institute's mission is educational outreach. "Everyone talks about it; we're really doing it!" says Hudson. The institute hired Neil Lamb from Emory University, who has developed a genetics course for middle schoolers that is used statewide, with a high school module following suit. A new conference center will handle demand for various adult education courses that are being offered.

A software company called Digital Radiance provides interactive teaching tools to students using video game technology. It was originally developed as another educational tool, but Hudson says a consultant got so excited by its potential that it was spun out as a new company.



Where bioscience companies have global reach

No wonder Georgia is hosting the 2009 BIO International Convention. In Georgia, world-class research, preeminent universities, and global accessibility accelerate almost 300 bioscience companies. Georgia is among the top 10 states for research and development, thanks to institutions like Emory University, Georgia Tech, and the University of Georgia. It's also home to the world's top public health agency, the U.S. Centers for Disease Control and Prevention.

For an up-close look at what Georgia has to offer, call 404.962.4040 to schedule your VIP site tour. Visit www.georgiabiosciences.com.

Visit us at the 2009 Bio International Convention in Atlanta at the Georgia Pavilion (#2817) and register to win a **KIA Soul!**



SPAIN

Spain Biotech: Leading the Way in Biotech Research

Spain has a long tradition of scientific excellence, particularly in life sciences. This is largely driven by a wealth of human capital with more than 9,000 researchers in universities, public research centers and hospitals that make up three-quarters of the Spanish scientists dedicated to biotechnology.

Excellent human capital combined with aggressive investment in biotech research and infrastructure has led to the creation of *bioclusters*; where regional governments such as those in Catalonia, Madrid, Valencia, Andalusia and the Basque country invest in coordinating public and private biology research and promoting the creation of knowledge-based companies.

Science Parks Promote Discovery

An important component of these *bioclusters* are the science parks that allow young companies to share sophisticated facilities, instruments and equipment to further their discoveries. Connecting these science parks with hospitals, universities and private companies creates a strong biotechnology sector within a region that attracts top talent as well as local and international biotech and pharmaceutical companies.

Today, more than 700 companies in Spain are involved in biotechnology related fields. More than 500 of which are R&D firms, and each year, 10 to 12 startups spin out of government institutions or large Spanish pharmaceutical companies.

Rich Pipeline of Products

The creation of *bioclusters* in the different regions has led to a robust pipeline of products and technology. A comprehensive portfolio of health care products issued by The Spanish Association of Bioenterprises (ASEBIO) includes 37 Spain-based biotechnology companies investing in 173 human drug projects; 40

new disease diagnosis and prognosis systems; and 12 animal health care products. Spanish companies feature more than 80 developments in clinical and preclinical phases of innovative molecules for insufficient illnesses treated. The first and most successful oncology product is Yondelis, PharmaMar's first product to go on the market.

Innovative Information Technology

The emphasis on biotechnology in Spain is also giving rise to breakthroughs in information technology specific to the industry. Bioalma, a company within the Bionostika Group, has created "the Google for printed biological data – a system of information retrieval that can search texts and understand the biological meaning of the written

word. Their free online search tool novoseek.com is on its way to being the world's leading biomedicine search system.

Search optimization also inspired the creation of Intelligent Pharma, a Barcelona-based start-up that offers the computerized system Helios that searches for molecules that match the function of compounds searchers plug into the database. The system is particularly helpful if a researcher has a natural compound that's difficult to synthesize and wants to identify a more chemically available substitute.

Data management is also improving in biotech thanks to Spanish innovation with companies such as NorayBio and Integromics. NorayBio helps companies design systems to better manage sample banks and is creating a visualization system for biomarkers, while Integromics offers a number of solutions for managing gene expression data.

These and the many other biotech companies and regional *bioclusters* are helping Spain grow as an epicenter of biotech research talent and innovation.

The emphasis on biotechnology in Spain is also giving rise to breakthroughs in information technology specific to the industry.

Barcelona Biomedical Research Park courtesy of PRBB.



Since 2000, the **biotech** industry in Spain
has grown at a rate of **25 percent** each year.



españa,
technology
for life.

Spain passion for business. solutions for living.



Spain's biotechnology sector is growing 10 percent faster than the U.S.* Driven by a culture, government and private sector that are open-minded and optimistic, Spain-based companies are emerging as leaders in this dynamic field.

Discover how creative solutions from Spain-based enterprises can revolutionize your biotechnology business. Visit spaintechnology.com/biotech.

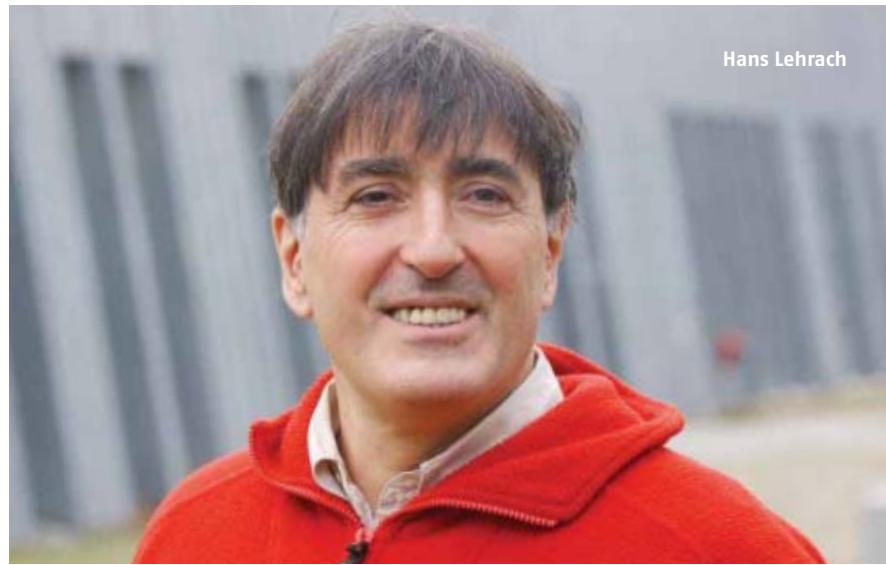
See us at
Booth 1405



spaintechnology.com/biotech

*According to
Genoma España

Computational Biology



Hans Lehrach

Hans Lehrach's Predictive Biology Philosophy

Alacris Pharmaceuticals, a new venture with George Church, applies modeling to individualized medicine.

BY KEVIN DAVIES

Hans Lehrach, one of Germany's foremost molecular biologists and a director at the Max Planck Institute for Molecular Genetics in Berlin since 1994, has frequently indulged his entrepreneurial leanings. In the early 1990s, he co-founded Sequana Therapeutics with other genomics rock stars such as Peter Goodfellow (see, "Genomics Provides the Kick Inside," *Bio•IT World*, Nov 2003), before its fortunes faded and its assets ended up at Celera. Another venture, GPC Biotech, did well until it suffered what the soft-spoken Lehrach calls "a somewhat catastrophic event" courtesy of the FDA in 2007.

Lehrach's forte is his gift for developing molecular biology tools and automation, from new gene cloning techniques to

chip-based platforms. His latest venture however, in collaboration with Harvard Medical School's George Church, stems from modeling programs he began writing on transatlantic flights ten years ago. The new company, [Alacris Pharmaceuticals](#), marries next-generation sequencing, systems biology, and computing horsepower, to turn complex biological data into a form that can be used by the researcher, the doctor, and the patient.

"Science is useful if it can predict things," says Lehrach. "Meteorologists are useful if they can predict the weather! Most of the time, taxpayers aren't interested in how many high-impact papers somebody publishes, but much more if he can predict something useful."

Lehrach and colleagues have been working on systems modeling approaches

to make predictions out of large amounts of data. Lehrach says there were three reasons to establish Alacris: modeling systems are improving, DNA sequencing is getting cheaper, and third, "GPC fired most of their researchers and many of their executives, and we had volunteers to set up a company!" Lehrach says jokingly.

Lehrach agrees that systems biology hasn't had a good rap in pharma circles. "Academic systems biology is very much single genes, single pathways," he says. "It's a great way to get papers published but not to make decisions about complex situations. Many people are just using it as a buzz word. One group says systems biology is nonsense and it shouldn't be funded, it's detracting from cell biology. Another group says systems biology is cell biology because cells are whole systems, so systems biology should fund my work!" Lehrach laughs. "The hard core of systems biology is to make predictions and predictive models."

Lehrach has known Church since George was a Harvard grad student in the mid 1980s. Lehrach admires Church's many ventures, noting: "Predictions of cancer drug response is a very low hanging fruit compared to what George is trying with the Personal Genome Project!" The other scientific co-founder of Alacris is Bernhard Hermann, one of Lehrach's former grad students, who is also a director at the Max Planck. "It's an enormous advantage to have him involved," says Lehrach. Hermann is a biologist who thinks about pathways, whereas Lehrach thinks more abstractly. On the business side, the company has a number of co-founding pharma executives and on the medical side is co-founded by top cancer doctors and researchers from the Charité Universitätsmedizin hospital in Berlin.

Computational Biology

Lehrach's modeling mission is to duplicate the biology of a cell or pathway in the computer. "Each object in biology corresponds to an object in the computer, which basically interacts with all the other objects," he explains. The starting point was a seminal cancer review paper published in *Cell* in 2000 by Doug Hanahan and the Whitehead Institute's

Bob Weinberg, Lehrach tried to set up a version of the cancer signaling pathways from the paper in the computer, setting up systems of differential equations that can be solved.

This is not a new idea; so why does it work? "It's not a problem to model differential equations if you know which constants to put in," he says. In biology, kinetic rate constants typically aren't known. So Lehrach takes each unspecified kinetic constant, and performs hundreds of modeling runs using different starting values drawn from probability distributions. "If everything agrees—if one result comes up—that's probably solid," he says.

The system, he says, must be tolerant against a lack of knowledge. "The problem up to now wasn't that we couldn't solve differential equations, it was that we didn't know how to handle uncertainty. Many people have tried to generate Boolean networks that pretend that the tumor is a Pentium chip with binary decisions, or basically tried to ignore the complexity

and uncertainty."

The most exciting model so far is that of the EGF receptor pathway, studied in the presence or absence of a downstream RAS mutation. "Using the model, we find that the RAS mutation increases cyclin levels just as much as EGF signaling. This predicts that EGFR antagonists will not have any effect in tumors with RAS mutations." This is indeed what has been found over the past 12 months by companies such as [Amgen](#) and [ImClone](#). (See "Amgen's Personalized Medicine Story," *Bio•IT World*, April 2008).

Lehrach calls his model "a virtual human, which oncologists can use to try a specific drug for a specific patient if you have the sequencing information." (See Box.) The first results are "extremely encouraging for predicting how patients react to existing therapies," he says.

This model is just the first pass. Lehrach is encouraged by the results obtained with just a few hundred runs, letting the computer run overnight in a research



George Church

institution with a small group. "Once you have 250 runs that give you the same qualitative result, then you can be confident. I'm sure we'll find better mathematical methods as time goes on to cut down on the computing time." Additional models are being developed for protein kinase inhibitors to explore how specific tumors will respond.

(CONTINUED ON PAGE 28)

CollabRx ONE, TREAT1000: the Cancer Connection

CollabRx was set up by Jay "Marty" Tenenbaum and Raphael Lehrer to slash the cost and time of developing drug therapies (see "Collaboration and the Long Tail of Disease," *Bio•IT World*, March 2009). Its new service, CollabRx ONE, applies those resources to help identify bespoke therapies for late-stage cancer patients. The goal of the service, which costs between \$50–100,000, is to provide a deep understanding of the individual patient's disease by marrying genomic and computational analysis, and match the aberrant target or pathway with a potential therapy.

"In every case we have actionable hypotheses that the doctors have not previously considered. It's not cheap, but we're working hard in order to be able to reduce costs... and we're actively seeking collaborations with major medical institutions," says Tenenbaum.

Lehrer, who is in charge of CollabRx ONE, has been a friend of Tenenbaum's for 20 years. After getting his PhD in physics from Harvard, Lehrer left research and moved into consulting and biotech, spending five years at Gene Logic working on toxicogenomics platforms and drug repositioning efforts, before finally joining forces with Tenenbaum.

CollabRx ONE has an informatics platform that builds tools for integrating a variety of data streams—gene expression, SNP analysis, copy number variant (CNV) analysis, sequence data—and potentially provides decision analysis. The small

informatics group in California is led by co-founder and chief technology officer Jeff Shrager (best known for writing an application called BioBike). Meanwhile, the wet-lab analysis is outsourced to CLIA-certified labs.

Lehrer says the innovations are chiefly in trying to distinguish signal from noise using so few samples. "Happily I'm a physicist," he says. "Hopefully you create ideas you happen to have training for!" The informatics group handles tasks such as pathway analysis (using Ingenuity's IPA) and data visualization, customizing ways of visualizing the data given the variety of data types.

Once the analysis is complete, CollabRx ONE staff meet with the patient's oncologist and discuss their findings, hopefully to advise on potential drugs or drug combinations, based on information on drugs that are either FDA approved or in clinical trials. Tenenbaum says his group is applying selective use of proteomics and "trying to understand the values of whole-genomes sequencing."

CollabRx has formed a joint project with Alacris Pharmaceuticals called TREAT1000, to add whole-genome sequencing to the CollabRx ONE offering. By sequencing 1000 genomes of cancer sufferers, TREAT1000 will not only provide potentially life-saving information about individual cancers, but create a compendium of cancer genome information that will inform

(CONTINUED ON PAGE 28)

Computational Biology

(CONTINUED FROM PAGE 27)

Lehrach is agnostic on how the sequence information is generated, although he hopes to use the Polonator, the low-price next-gen sequencer based on Church lab technology. When we spoke, however, he had shipped the instrument back to the U.S. for repairs and upgrades. For now, the focus is genomic data, but Lehrach will take any form of data: copy number variants (CNVs, low coverage sequencing of the genome), exome sequencing for the patient, tumor and if possible, the tumor stem cells; and transcriptome data to monitor epigenetic effects. "Proteomic data would be fantastic but it's still a challenge. We're working on it," he says.

Safety Switch

Ideally, Lehrach says, "we would tell patients not only which drugs won't work

but also, at some point, which drugs will work." Funding for individualized medicine—sequencing patient tumor genomes, predicting drug response, and developing personalized drugs—could come from the patients themselves or insurance companies.

Another goal is to improve drug safety. Lehrach says the EGFR antagonist trials only worked because they identified a subgroup of non-responders with the RAS mutation. "If we're able to predict the number of real responders, you can run a trial with ten patients! The same P value can be achieved at much lower cost and much shorter time." Sponsors would see their market shrink, of course, but the reduced trial costs and increased lifetime of the patent (because of earlier approval) would more than compensate. In short, "the more work we can move to the com-

puter, the better things are going to go."

Alacris Pharmaceuticals would eventually like to engage with big pharma, but many pharma firms come from the single gene tradition, says Lehrach, so it's hard to convince them of the value or validity of this approach. Some are interested, however, despite having "been burned by a lot of hot air." It wouldn't be difficult to incorporate pharmacogenomic information from the cytochrome P450 genes from the same patient, "so there's a hell of a lot we can predict from sequence data."

The name "Alacris" suggests promptness, but Lehrach is gearing up operations methodically. The company was founded last year—to say it has "launched" would be an overstatement, Lehrach admits. So how does he go about raising money in this economy?

Lehrach laughs: "Any suggestions?" •

(CONTINUED FROM PAGE 27)

future research and treatment.

What the Data Say

"We look for what the data are telling us, and how that matches the therapies," says Lehrer. Leading the analysis is Bob Coopersmith, a former Gene Logic colleague. "We push back and forth conclusions and alternate explanations, discuss discrepant observations," says Lehrer.

Although the project is in its early days, Lehrer says progress based on the first half-a-dozen patients is extremely promising. "It's important for us to connect with the oncologist and that the oncologist buys into what we're doing." In nearly every case, "the oncologist has been pretty excited," although it's too early to predict how that might translate into clinical success.

The work bears an emotional toll. One patient, thought to have 12 months to live, died in a matter of weeks before the team could implement their findings. In other cases, there are signs of a partial, but only partial, response to drug. Nevertheless, for Lehrer and Tenenbaum, that's encouraging. In the case of a lung cancer patient on Avastin, CollabRx analysis revealed the likely involvement of two different key pathways. "The second was going untreated, and that suggested a combination of drugs, or one drug that could hit both pathways." But the oncologists must follow the standard of care, which typically means changing the drug regimen only after the patient becomes fully resistant.

More Online

For further information on CollabRx ONE, see the video interview with Marty Tenenbaum: www.bio-itworld.com/lsw/jtenenbaum.



The Treat1000 team including Lehrach and Church (front row, left and center), and Coopersmith, Lehrer, and Tenenbaum (back row from right).

The recommended drugs are likely to be off label. "They may or may not be cancer drugs. They may or may not be generic. They may or may not be reimbursed," says Lehrer. "Oncologists may have tried it, but the reaction is often, 'I wouldn't have considered using that kind of drug, but based on what you're showing me, that makes sense.'"

Lehrer stresses that the results are always communicated to the patient's oncologist, "because they are the ones who need to decide whether what we've found is something that should be tried, or if additional studies should be done. Depending on circumstances, it is valuable to have the patient there as well."

CollabRx ONE is continuing to evaluate new technologies, studying whether a potential data source can add value to an actionable hypothesis. Lehrer says: "If it is adding value, are previous sources now redundant? Obviously it's not worth doing SNP analysis if you're doing whole-genome sequencing." K.D.

Bioalma Launches novo|seek PubMed Search Tool

Intelligent search ranks documents according to relevance, and researchers by expertise.

BY KEVIN DAVIES

Bioalma, the Spanish biomedical IT/text-mining company, has launched a free search tool for the PubMed literature database called *novo|seek*, which the company claims provides intelligent search functionality to help life scientists guide and refine their searches of the biomedical literature.

The company calls *novo|seek* "a dynamic information extraction system" for searching biomedical records in repositories, particularly PubMed. *Novo|seek* indexes the biomedical literature in PubMed and enables researchers to find relevant results efficiently by using external sources of data and contextual term information. The tool provides familiar chronological listings of search results, but a sidebar presents a series of additional related terms based on relevancy, allowing researchers to drill down and refine additional queries.

Bioalma CEO Juan Carlos del Castillo said: "Novo|seek leverages the same Medline data as PubMed, but our search capabilities are second to none and yield better, more relevant results much faster and easier and don't require a lot of training." *Novo|seek* ranks documents according to their relevance to identify the most interesting information through a recognition of key biomedical concepts. It also ranks authors, letting users know who the main expert is in relation to the query searched.

Search and Deploy

One scientist familiar with *novo|seek* is Reinhard Schneider, team leader in Data Integration and Knowledge Management at the European Molecular Biology Laboratory in Heidelberg. "I think it's a very nice tool for researchers, which links the web site with the Deep Web," said

Schneider. "I think it will change the way we work on the web in the future."

Schneider says *novo|seek* is a powerful search tool that provides depth of information "typically only found in large text mining machines. With the easy-to-use interface and the ability to filter results in a number of ways the system enables me to drill down to the most relevant information very fast and efficiently."

Schneider, who has worked with the tool for the past few months, says Bioalma has developed dictionaries of proteins, diseases, chemical names applied to the PubMed abstracts. He calls *novo|seek* "a tool that allows you to find the most relevant information and put that into context. If you look for diseases, it will give you feedback on what kind of proteins are involved in the disease, what kind of chemicals, and so on. Instead of Googling for hours when you find a new term, you

4 I think it's a very nice tool for researchers, which links the web site with the Deep Web. I think it will change the way we work on the web in the future."

Reinhard Schneider, European Molecular Biology Laboratory

get a very fast overview of what is known about the topic."

He gives Huntington's disease as an example. The system returns the standard Medline results, but the sidebar lists a panel of concepts, such as diseases and syndromes, pharmacological substances, proteins, chemical substances, and organisms. "If you click on these terms, you jump from your initial search term to a protein or a gene name. So it's very easy to get an overview of what is known about the disease," says Schneider.

Ramón Alonso-Allende, business development director, says Bioalma has been perfecting its information extraction technology for the past five years. "We use statistics, natural language processing and context information. We are able to analyze the literature and extract the knowledge contained [therein]—we're not just doing pattern matching."

Bioalma downloads and indexes 18 million documents in Medline each day. That information is then put into the company's own database using the open-source Lucene search engine library. "We don't use NCBI resources to pull data on the fly," said Alonso-Allende.

General manager Luis Cornide explained that unlike the AKS2 product, which operates in a standard license model, *novo|seek* operates under a different model. "Our goal is to be there every day with the scientist," Cornide said. "Novo|seek is designed and conceived to be used by the whole biomedical community, accessed by lots of people to obtain results in Google standard time."

While Bioalma believes that *novo|seek* will help introduce its other products to a broader audience, it also hopes to generate revenue by selling online advertising through Google ads, targeting companies selling reagents or equipment. Bioalma has not held discussions with NCBI about the tool. Alonso-Allende admits that, "Trying to get some of their pie is very challenging. PubMed is a great standard, but it lacks the intelligence that is out there that we can provide with our technology."

The latest release of *novo|seek* includes additional resources such as grant information and full-text search based on open access and PubMed Central sources. •

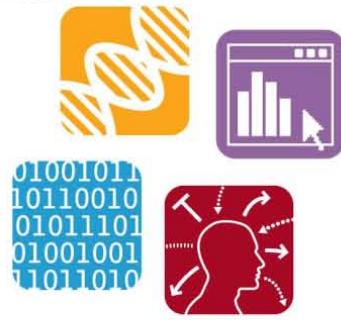
Register by 26 June and Save up to €260!



EUROPE

Bio-IT World CONFERENCE & EXPO '09

6–8 October 2009 • Exhibition Grounds • Hannover, Germany



Enabling Technology. Leveraging Data. Transforming Medicine.

CONFERENCE TRACKS:

Track 1: IT Hardware for the Life Sciences

Track 2: Bioinformatics for Genomics

Track 3: IT Software for the Life Sciences

Track 4: Data Integration and Knowledge Management

Event Features:

- Hear over 70 innovative technology and scientific presentations arranged in four tracks
- Networking opportunities to meet and greet other life science and IT professionals from around the world
- Participate in the poster competition
- View novel technologies and solutions from 850 companies in the expansive exhibit hall
- And much more!

Held in Conjunction with:



Europe's No. 1
Event in Biotechnology
and Life Sciences

Corporate Sponsor :



Official Publication:

Bio-IT World

Organized & Managed by:

Cambridge Healthtech Institute
250 First Avenue, Suite 300, Needham, MA 02494
Phone: 781-972-5400 • Fax: 781-972-5425
Toll-free in the U.S. 888-999-6288



Bio-ITWorldExpoEurope.com

DISTINGUISHED FACULTY

- Chris Asakiewicz, Ph.D., Affiliate Professor of Information Management, Wesley J. Howe School of Technology Management, Stevens Institute of Technology
- Herbert Auer, M.Sci., Director, Functional Genomics Core, Institute for Research in Biomedicine, Parc Científic de Barcelona
- Shiva Ayyadurai, Ph.D., Fulbright Scholar & Faculty Lecturer, M.I.T.; Executive Director, International Center for Integrative Systems Research
- Andrew Berridge, B.Sc., Delivery Advisor, Informatics, Pfizer
- Bernd Bischl, Fakultät Statistik, Technische Universität Dortmund
- Michał Blażejczyk, M.Sc., Lead Programmer, Bioinformatics, Génome Québec and Montreal Heart Institute Pharmacogenomics Centre
- Lars Carlsson, Ph.D., Global Safety Assessment, AstraZeneca R&D, Mölndal, Sweden
- Karim Chine, Software Architect and Coordinator, Biocep
- Jacob DeVlieg, Ph.D., Global Head, Molecular Design & Informatics, Schering Plough
- Christian Herhaus, Ph.D., NCE Technologies, Merck KGaA
- Andrew L. Hopkins, D.Phil., FRSC, SULSA Research Professor, Translational Biology & Chair, Medicinal Informatics Division, University of Dundee
- Andreas Keller, M.Sc., Head of Analytical Service, febit biomed
- Janice Kielbassa, Ph.D., Researcher, Biometry and Evolutionary Biology Laboratory, University Claude Bernard
- Jan Korbel, Ph.D., Group Leader, Gene Expression, European Molecular Biology Laboratory
- Jerry Lanfear, Ph.D., Head, Data Support and Management, Research CoEs, Pfizer
- Thomas Manke, Ph.D., Researcher, Computational Biology, Max Planck Institute for Molecular Genetics
- M. Scott Marshall, Ph.D., W3C HCLS IG co-chair, Informatics Institute, University of Amsterdam
- Sergey Nejentsev, M.D., Ph.D., Royal Society Research Fellow, Department of Medicine, University of Cambridge
- Ingela Nyström, Ph.D., Centre for Image Analysis, Director of UPPMAX, Uppsala University
- John Overington, Ph.D., Team Leader, EMBL-EBI
- Julen Oyarzabal, Ph.D., Head, Drug Discovery Informatics Section, Spanish National Cancer Research Centre (CNIO)
- Tommaso Piazza, Eng.D., Chief Information Officer, Mediterranean Institute for Transplantation and Advanced Specialized Therapies (ISMETT)
- Josef Scheiber, Ph.D., Research Investigator/Knowledge Engineering Consultant, NITAS/TMS, Novartis Pharma AG
- Markus Schirmer, M.D., M.Sc., Junior Scientist, Clinical Pharmacology, Georg-August-University
- Reinhard Schneider, Ph.D., Team Leader, Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL)
- Ola Spjuth, M.Sc., Bioclipse Coordinator, Uppsala University
- Mats Sundgren, Ph.D., Principal Scientist, Global Clinical Development, AstraZeneca R&D
- Fabio Triolo, D.d.R., M.Phil., Ph.D., Technical Director, Regenerative Medicine and Cell Therapy Unit, Mediterranean Institute for Transplantation and Advanced Specialized Therapies (ISMETT)
- Jan van Oeveren, Ph.D., Biostatistician, Bioinformatics Department, Keygene NV
- Tibor Van Rooij, B.Sc., Bioinformatics Director, Génome Québec & Montreal Heart Institute Pharmacogenomics Centre
- Jörg Kurt Wegner, Ph.D., Senior Scientist, Bioinformatics, Tibotec (J&J, Belgium); Blogger, Mining Drug Space; Project Administrator, Open Source Development

**For more speaker additions, please visit
Bio-ITWorldExpoEurope.com**

Register by 26 June and
Save up to €260!

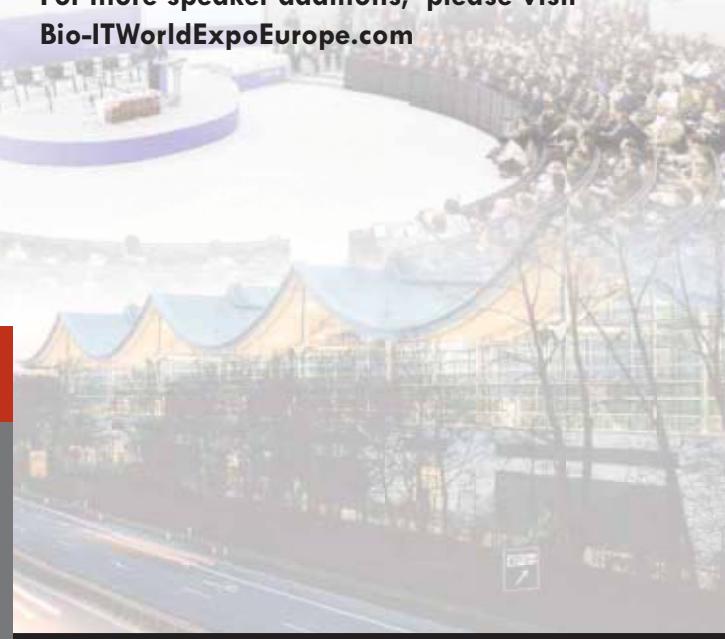
To Register...

Web: bio-itworldexpoeurope.com

Phone: 781-972-5400 or toll-free in the U.S. 888-999-6288

**Please mention keycode
U02 when registering!**

Bio-ITWorldExpoEurope.com



Computational Development

Novartis Savors Early Modeling Success

Under Don Stanski, Novartis' expanding Modeling & Simulation franchise is paying dividends.

BY JOHN RUSSELL

CAMBRIDGE, MA.—Modeling and Simulation (M&S) has a complex history in the life sciences. Big pharma and small biotechs alike have struggled to find the right recipe. Like a good wine, having the right basic ingredients—management buy-in, modeling talent, sufficient resources, suitable problems—is necessary but not enough for success. A certain amount of vintner's craft and aging has to occur and quite a few companies have gulped some pretty rancid stuff along the way.

Novartis has evidently found ingredients and conditions that bode well. Some three years ago, Donald Stanski convinced Novartis that M&S would never fulfill its potential embedded in the biostatistics department. He argued it should be set apart, closely aligned with Novartis' therapeutic franchises, enjoy the right reporting structure, and be given enough resources and time to prove itself. Modeling scientists would sell themselves to the drug development project teams and sink or swim based on results, deliverables, and value, argued Stanski.

Several key personnel were already in place, including Gabriel Helmlinger, director and global head, M&S-Biology. And senior management was ready to buy in. Trevor Mundel, then head of Translational Medicine, is now Novartis' global head of drug development, based in Basel. He's a physician scientist who did graduate work in the applied mathematics of dynamical systems.

"I've always seen modeling and simulation as a key technology that really hasn't been as widely deployed or as effectively deployed as it could be," says Mundel, his voice booming over the transatlantic speakerphone. "We're now having the



Don Stanski sends an M&S wake-up call.

chance to work with Don's group and to give it its proper place in the context of development. We are starting to see some real traction."

Not surprisingly, Mundel is focused on the phase II problem—attrition, duration, and cost—and "how one de-risks the process with proof of principle efforts. Given that proof of principle gives you a pretty small data set, how can you play that up and extract more information so you don't have to spend excessive time in phase two? We bring in modeling and simulation as one aspect of it."

Stanski's solutions seem to be paying off. He joined as VP and global head of M&S, expanding his team from 30 to 50 scientists the first year. He now has approval to add another 15 staff. The M&S group has even attracted funding for some additional FTEs from Novartis business units such as Oncology, Sandoz, and Animal Health.

"One of the coolest problems we've been involved with was with Animal

Health," says Stanski playfully. "It was around dosing for cows. They pour the medication on the backs of the cows and the cows self-dose by licking each other's backs. The question was around the dose to the cows and the resulting kinetics in the animals."

Broad Scope

The sheer breadth in scope of Novartis' M&S department activities is impressive. Biological modeling (pathways, mechanisms, drug-disease), pharmacological modeling (PK/PD and trial design), biophysical modeling (organ- and tissue-level models), and decision analysis are all part of the mix.

The scope of questions tackled is equally expansive and generally familiar: appropriate biomarker selection and measurement strategy, dosing guidance, candidate and target selection, clinical study design, and competitive analysis. Though impact is difficult to measure, Brian Stoll, a senior expert modeler, says, "Today it's unlikely that a new biologic candidate would make it into Novartis' late stage pipeline without contributions from M&S."

The M&S department's new digs in Cambridge, Mass., where it is co-located with the Novartis Translational Sciences group, adjacent to the former Necco candy factory (see "The Sweet Hereafter," *Bio•IT World*, July 2004) are distinctly high-tech and sleek. No dingy, windowless cubicles for this crowd.

It also helps that Stanski has been a long-time evangelist for modeling approaches. Trained as an anesthesiologist, he is a clinical pharmacologist with stints in industry (VP at Pharsight), academia (chair, department of anesthesia, Stanford), and the FDA (advisor for the director, CDER).

"Most companies create modeling silos by distributing the modeling talent throughout the company," he says. But separating biological pathway modeling from PK/PD modeling or statistics is usually a mistake. All of those disciplines need to be brought to bear on problems and the cross-fertilization is essential

and opportunistic. “I had never worked with this [biological pathways modeling] before,” says Stanski. “Gabriel and his colleagues represented learning for me. When I saw what they could do and how you could integrate it, to me that represented one of the most interesting high value opportunities.”

Stanski explains the Novartis M&S team is organized loosely around “three dimensions”: modeling skills, therapeutic drug development, and software working platforms. Modeling skills embraces biological pathways, Bayesian statistics, and trial design, among others.

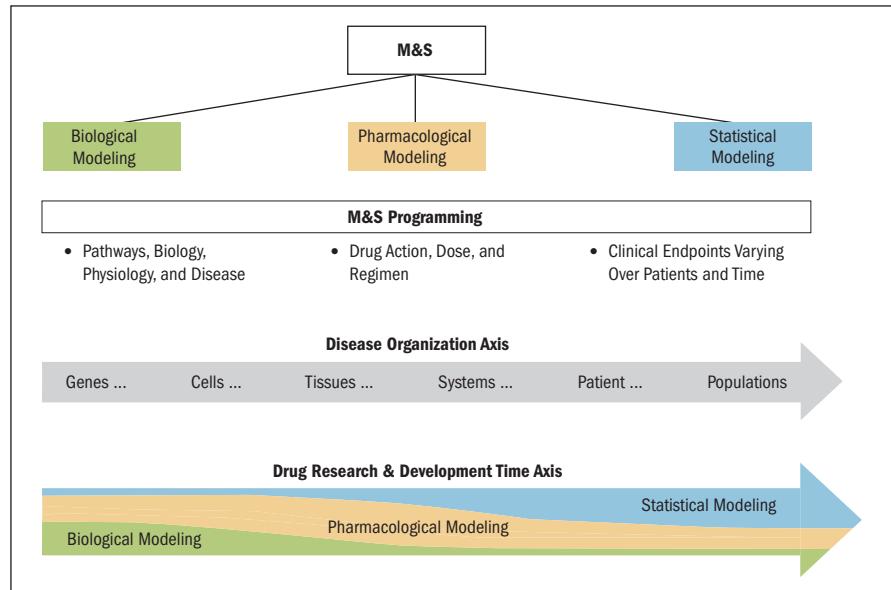
The department has organized “clusters” to mirror the Novartis therapeutic franchise drug development approach. The heads of each M&S modeling cluster work closely with their drug development franchise counterparts to identify opportunities to add value.

“We align ourselves based upon how Trevor has set up the therapeutic drug development focus of the company. We [M&S] call them clusters to be a little different,” says Stanski. “These are the basic therapeutic franchises in Novartis. Oncology is a separate business unit. But we do the modeling for them and they provide funding for FTEs.”

Platform Boots

The organizing principle, Platforms, refers to re-useable modeling tools that address a recurring problem in Novartis M&S jargon. “We have different types of modeling platforms,” explains Helmlinger. “Some are very methodology-driven. For example, look at data mining. There are many algorithms out there—classifiers, search algorithms, inference algorithms. If we get a new question around data mining, we might have to search the whole methodology background of the algorithms to see which one we should try. But we have organized this into a library form, so we know which types of algorithms we can use and which are software-ready to apply to a particular problem.”

Helmlinger indicates, “There are also safety (cardiac safety modeling is one) as well as drug-disease modeling platforms. Take diabetes, we have addressed many diabetes problems through our projects using PK/PD modeling, so there’s a whole



M&S integrates biology, pharmacology, and programming across the pipeline.

range of PK/PD models with different focuses. We capture all of those in a diabetes drug-disease modeling platform. Furthermore we can then integrate these models maybe to gather the molecular pathways that we know of and how we think they can be networked into one, let's say, higher level model. So that's the goal.”

Platforms represent both institutional memory and practical tools. Currently the M&S team is the prime user but there is no reason, and every hope, that a wider Novartis constituency will come to use the platforms. The cardiac safety platform has been shared with Novartis basic researchers who evaluate the electrophysiology of all compounds that will be given to humans.

A good deal of IT support is obviously required for M&S. Stanski says they work closely with the Novartis IT services but we need to also make our own major contributions. “Our IT needs are specific to what we do. It's been time-consuming, extensive, and painful, but it's finally starting to work. We have had 2 to 3 FTEs working on IT issues for the past several years. That's expensive for a group of 50, but you know what, there's been no choice.”

Winning acceptance from the therapeutic drug development franchises is the key to sustaining Stanski's vision. As the saying goes, nothing succeeds like success. Stoll offers two examples of dis-

covery-related projects, which he classifies as “mechanistic modeling”—modeling where “we try to capture the underlying biophysics and biology and for which the parameters have a meaningful physical interpretation.”

One example involves the rational selection of antibody drug candidates most likely to succeed in the clinic. In this case, there is a competitor already on the market that hits a different target in the pathway. Leveraging knowledge about the disease and some clinical data, “we develop a model of cell signaling pathways in which an ordinary differential equation describes the change in each molecular species with respect to time,” indicates Stoll. “We used the model to identify candidates that would not only be successful but that would also be well positioned against our competitor.”

The uniformity in the pharmacokinetics of certain antibodies along with an understanding of the underlying biology enables “model prediction of relationships such as dose-response curves very early in discovery, even before certain pre-clinical studies are conducted.”

The model provides quantitative evidence that their target has potential and helped narrow the pool of candidates to be evaluated further. Moreover, Stoll says, “We can further compare the cost of time spent further maturing the antibody affin-

Computational Development

ity and compare it with the savings in cost of goods associated with administering a lower dose because you have a higher affinity. This is an example of tradeoff calculations we can make to facilitate more informed decision-making." The franchise development project team makes the final decision, using the M&S information along with multiple other factors.

Stoll presents another more complex example involving a signaling pathway with many non-linear characteristics. Among other things, the purpose of the modeling was to understand the activation dynamics of biomarkers of response. Both modeling efforts are *in silico* efforts validated using experimental data.

In many ways, these modeling analyses integrate data and present them in quantitative ways to assist decision making for the drug development team. This is commonplace in other industries in which the underlying phenomena are better understood. But in the pharma/

biotech industry, getting from the data to the models and to their revealing predictions can be an arduous process.

Spinal Tap

It is easy to forget how important macroscopic modeling can be. Helminger showed Novartis work modeling the spinal column and the flow of cerebrospinal fluid. One impressive example involved the application of an antibody into the cerebrospinal fluid to treat spinal cord injury. Despite good efficacy data in animal models, the big question was: how to scale it up to humans? In the classical PK/PD world, Helminger explains, some methods are quite simple to scale up to humans. Dealing with the spinal cord, however, presented an entirely different geometry, with complications surrounding fluid dynamics and kinetic sampling.

Like designing an airplane, Helminger's group built part of the human body on the computer—the spinal cord and

surrounding tissues—using a biophysical modeling platform. From an engineering standpoint, the challenge was "injecting a large molecular weight antibody into this unique, anatomically specific space, the intrathecal region." Depending on the site of injection, will the drug get to the site of action? Variables include the syringe type, its orientation, location, rate of injection, and the patient's orientation (standing or laying down).

All of this can be entered in such a biophysical model. Helminger says: "Knowing from preclinical data that the antibody is efficacious, we just want to make sure it gets to the site of injury in patients with a spinal cord injury."

Another impressive example was senior modeler Kai Wu's description of an integrated drug efficacy and safety modeling analysis to determine the optimal dosing regimen. The Novartis M&S team seems bent on leaving no stone unturned with regard to how M&S might be applied



Visit www.eCliniqua.com today!
Subscribe to the eCliniqua eNewsletter today!



CAMBRIDGE HEALTHTECH MEDIA GROUP

Your Connection to Clinical Research

Introducing *eCliniqua* – a new web site focusing on innovative management in clinical trials covering specific topic areas including:

- Trial design
- Data capture
- Regulatory & safety
- Market insights
- Technology solutions
- Other clinical-related critical topics

The web site is the *expanded face* of the *eCliniqua* eNewsletter, one of the most respected sources for news covering clinical research and trials. Both *eCliniqua* mediums provide life science professionals with critical and dynamic coverage on the clinical research enterprise.

Stay informed. Remain competitive.

Bio-ITWorld.com • CHIMediaGroup.com

eCliniqua—Innovative management in clinical trials

to enhance decision making and progress toward corporate goals.

One unusual effort is decision analysis, led in Cambridge by expert modeler James Dunyak, a former mathematics professor at Baylor. His work involves examining all of Novartis' drug development practices and the associated decision making processes.

"Our vision of decision analysis, actually, is that it should be used at all levels of the company," explains Dunyak. "I'm doing some work in early development on how we treat patients in clinical trials in different countries. There are all kinds of issues associated with manufacturing and formulation and there are almost always unsolved technical issues, not just tactical issues of where are you going to make it and what are the raw ingredients, but there's usually formulation issues. For many projects I've ever been on, the problem has been described as less soluble than brick dust."

More prosaically, he presents work on a modeling project in which the original strategy (presented by the franchise development project) would have introduced delay without achieving substantial benefit; the strategy was subsequently changed because of his analysis.

Stanski says, "Our challenge is to find areas where we can mix decision analysis and economic modeling, to the drug development. I can't say we've got it all figured out. Some of it is opportunistic. You know, Jim has built credibility. Now they're coming back for more, which can create a capacity issue. This is early yet. We're six months to a year into this. But this isn't where we started. We started with Trevor on a proof of concept and drug-disease model and did that for 2 to 3 years."

So far, Stanski's commitment to keeping modeling scientists together, aligning efforts directly with the therapeutic franchises, and ensuring that modeling di-

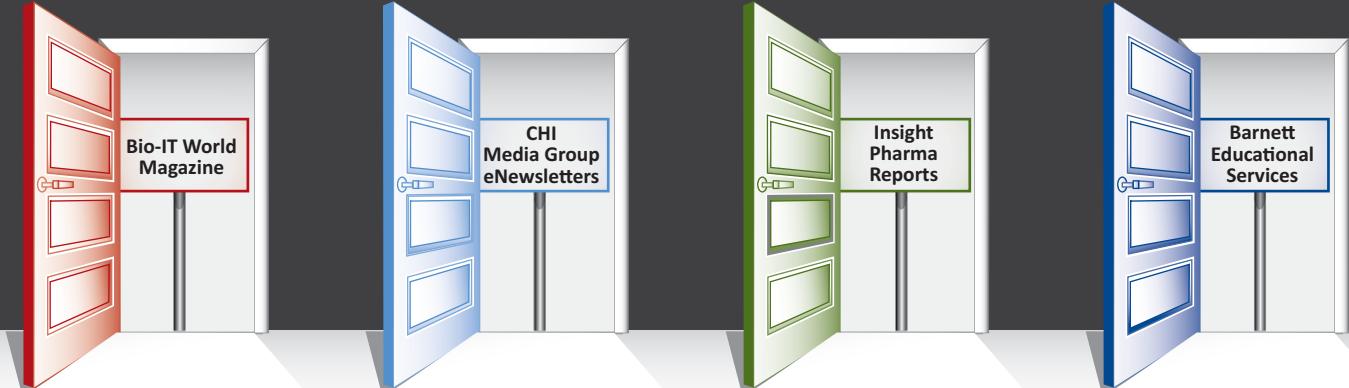
rectly supports decision-making activities rather than serves a broadly descriptive function, is paying off. Mundel's goal, after all, has little to do with aesthetically pleasing graphs and charts, and everything to do with accelerating and improving drug development.

The group keeps close tabs on itself. After every project, Stanski says, "We do a specific interaction interview with the key decision makers on what was the value of the modeling efforts. We also have our own software to keep track of projects and outcomes. When a project is done, there will be a value statement illustrating the impact. It's very tangible. We make a proactive effort to capture the value of everything we do," says Stanski.

What portions of this formula are transferable to other companies is far from clear. But Stanski's hopes for M&S at Novartis are ambitious. For now, the Novartis M&S "vinification process" is maturing nicely. •

Open Doors to Enhance Your Knowledge

 Cambridge Healthtech Institute
Your Life Science Network



Cambridge Healthtech Institute (CHI) offers a suite of published resources focusing on the life sciences industry.

- **Bio-IT World Magazine:** Covers critical areas of computational biology and development, clinical research, and IT informatics.
www.bio-itworld.com
- **CHI Media Group eNewsletters:** Three topic-specific eNewsletters covering clinical trials, predictive biomedicine informatics, and technology driving drug discovery.
eCliniqua • Predictive Biomedicine • Bio-IT World Weekly Update
www.chimediatgroup.com
- **Insight Pharma Reports:** Series of well-written, concise reports evaluating the salient issues in pharmaceutical technology, business, and therapeutic markets.
www.insightpharmareports.com
- **Barnett Educational Services:** Reference manuals focusing on drug development, R&D productivity, regulations, and industry trends.
www.barnettinternational.com

Published Resources for the Life Sciences

Enhance your knowledge today by subscribing or purchasing a publication that meets your business and professional needs. Visit the web sites for more information.

Feature

Learning to Embrace the eCTD

Sponsors, CROs, and industry providers learn the pros and cons of switching to the electronic common technical document.

By Ann Neuer

CONSIDER THE ENTICING PROSPECT of eliminating the blizzard of paper associated with regulatory submissions through a process in which applications are built once from the beginning, and then managed throughout their lifecycle. Imagine enhancing the quality and speed of those submissions, while establishing a history of the submission in an easily searchable form.

“Anything is better than paper,” says Nancy Smerkanich, VP Global Regulatory Affairs for **Octagon Research**, who sums up a widely held opinion in the industry. “Just being able to pull up a 3-year-old investigational new drug (IND) application and look at it in a current view as well as a historical view with a couple of clicks is an enormous advance over spending hours in a file room looking through boxes, pulling out volume after volume—in some cases, hundreds of volumes.”

The key to this significant change is the electronic common technical document (eCTD), which represents a major advance in the march toward electronic regulatory submissions across the globe. Put forth by the International Conference on Harmonisation (ICH) back in 2003, the eCTD specifies how electronic submissions are to be created, reviewed, and archived.

Following the 2008 mandate by the Center for Drug Evaluation and Research (CDER), a division of FDA, *Bio-IT World* set out to gauge the reaction of the major stakeholders—pharmaceutical companies, government, and industry providers—to the rising tide of eCTD adoption, and assess the benefits and challenges in making eCTD adoption a successful effort. Users are finding that creating the infrastructure to participate in the eCTD world is complicated, requiring a commitment to process change and long-range strategic planning. Regulatory bodies are seeing frequent errors in submissions that can hinder and possibly delay review, although many of these are preventable.

Looking ahead to future trends, industry insiders anticipate an impact of the Regulated Product Submission standard (RPS) on the eCTD, just as it is gaining traction. RPS is a Health Level 7 standard designed to facilitate processing and review of all regulated products, and within several years, eCTD is expected to be folded into that standard.

On the March

Long burdened with volumes, even truck-loads, of paper submissions that must be created, shipped, reviewed and archived, the drug industry is actively embracing the eCTD, signaling a new paper-free era for the regulatory submission process. North America, Europe, and Japan are experiencing continual growth, and other parts of the world, notably Australia,

Growth in eCTD Submissions to CDER*

	APPLICATIONS		SUBMISSIONS	
	Oct 2006	Oct 2008	Oct 2006	Oct 2008
IND	158	1,310	2,304	22,560
NDA	105	1,115	1,685	11,531
ANDA	76	1,296	318	4,857
BLA	31	89	859	3,685
DMF	20	241	23	434
TOTAL	390	4,051	5,189	43,067

* numbers are cumulative; "Submissions" refers to any individual item sent to a regulatory agency.

Source: CDER

which issued a draft guidance in January, are hoping to follow suit.

On January 1, 2008, CDER mandated that electronic submissions had to use the eCTD format. That decision sparked a sharp jump in the cumulative number of eCTD submissions, which tripled in a single year, from nearly 15,000 (October 2007) to more than 43,000 by October 2008. The number of total applications soared tenfold over the two years from October 2006 (390) to October 2008 (4,051). An application is defined as an investigational new drug application (IND), a new drug application (NDA), an amended NDA (ANDA), a biologic license application (BLA), or a drug master file (DMF).

Europe is also seeing strong growth in eCTD submissions in advance of the European Medicine's Agency's (EMEA) recommendation for eCTD-format elec-

tronic-only submissions starting July 1. According to a survey conducted by the Telematics Implementation Group for Electronic Submission, a group of representatives from European Union National Competent Authorities and EMEA, the number of eCTD submissions spiked almost 50% in a six-month period spanning 2007 and 2008. The volume of eCTD submissions grew from 1,435 in the second half of 2007 to 2,122 by the end of the first half of 2008.

By January 2010, the agency will mandate eCTD format for all electronic submissions using the Centralised Procedure, which is a method for medicinal products seeking approval for use in all EU countries, and applies to all applications and all submission types.

Pharma sponsors and outsourced providers are driving the expanding eCTD marketplace. Terri Booth-Genthe, senior director of global regulatory submission management for [Wyeth Research](#), says the company was an early eCTD adopter, dating back to a 2003 submission to FDA. Since then, Wyeth has submitted nearly 5,000 sequences, which are multiple submissions in support of a drug in clinical trials, mostly to FDA, but also to EMEA and Health Canada. Last year, Wyeth submitted a total of 2079 eCTDs, a 6% increase over the previous year.

Currently, Wyeth submits everything in eCTD that FDA is able to accept. Booth-Genthe explains that in 2007, Wyeth set about converting all of its active NDAs, INDs, DMFs and BLAs into the eCTD format. "By the end of 2007, all of our active U.S. files were in eCTD." The following year, Wyeth converted all of its

“Anything
is better than
paper.”

Nancy Smerkanich,
Octagon Research



Feature

Elementary eCTD

Here are a few basic facts about the eCTD structure to shed light on why and how it holds promise to foster significant change. The eCTD specification, developed by the M2 Expert Working Group (EWG) within the International Conference on Harmonisation (ICH), defines the criteria for the overall architecture of how the eCTD is to be structured so electronic submissions are technically valid.

In its most basic form, the eCTD is composed of three elements: a directory structure; an Extensible Markup Language (XML) backbone that creates links to leaf files; and content files. Essentially, it is the electronic version of its predecessor, the common technical document. The XML backbone enables management of metadata for the entire submission and for each document within the submission. It also provides a table of contents (TOC) and navigation aids to the individual files.

Reports and forms use a PDF format, and datasets are transported using SAS XPORT files.

A submission is organized using five modules (with appropriate navigational links):

- Module 1 – Administrative information and prescribing information unique to each region
- Module 2 – Summaries
- Module 3 – Quality
- Module 4 – Nonclinical study reports
- Module 5 – Clinical study reports

Each document should contain bookmarks and hypertext links from the TOC to all tables, figures, publications, and appendices. To enable navigation, documents should be generated from electronic source documents and not from scanned material.

European marketing authorisation applications (MAA) to eCTD format. That effort is paying dividends. Currently, there is just a 1-4 week lag between Wyeth's U.S. and EU submissions; prior to the eCTD initiative, Booth-Genthe says the gap between submissions was often 6-12 weeks.

GlaxoSmithKline has also embraced the eCTD, and prepares them using three different publishing tools—one for the United States, one for Europe and another for Canada and Japan—but is in the process of transitioning to a single global tool. Andrew Marr, Director Global eRegulatory Development, comments that in the U.S., over the past 18 months, Glaxo has transitioned all of its NDAs to the eCTD, bringing the total to approximately 140. In Europe, the company has submitted 28 MAAs in eCTD format to the European Medicines Agency through the Centralised Procedure. The company's current focus is on expanding the use of the eCTD to include the IND process. "We'll be piloting some INDs into CDER, trying to get the process right. We have already submitted a handful of eINDs into the Center for Biologics Evaluation and Research (CBER)," says Marr.

Small to mid-size players that lack the volume or infrastructure to justify developing internal capability often outsource this function to providers with established expertise in eCTD preparation and submission. Nancy Smerkanich

of Octagon Research says the company's eCTD volume has tripled since 2003-2004. "We have 80 active eCTDs that we have filed and maintained, and we recently celebrated our 1000th sequence going through the electronic submissions gateway at FDA. The majority of those 80 applications are for small pharma or those in the biotech area."

Similarly, **Apyx** is seeing a significant uptick in its eCTD business. According to president and CEO Ken VanLuvanee, eCTD represented less than 20% of revenues a couple of years ago, but that has now risen to 75% to 80%. "In approximately three years, when Regulated Product Submission (RPS) becomes a consideration, Apyx anticipates that 65% to 70% of company business will be electronic publishing, and of that, 90% will be in eCTD format," VanLuvanee explains. Approximately 70% of Apyx's eCTD submissions go to FDA, 25% to EMEA, and 5% to Health Canada.

Benefits Come With Challenges

Wyeth's Booth-Genthe sees the eCTD as providing major benefits beyond its role as an electronic submission tool. The information it contains can also bring substantial efficiencies to the internal organization. "More of our authors and regulatory liaisons are depending on eCTD files to give them background and they are starting to rely on these files as a

reference," she says. "They can easily go back and review individual components and their lifecycle. This adds significant value as many people in the quality organizations can now easily see changes in specifications we've made over time."

But reaching the point that an eCTD becomes integral to the operation—both as a submissions method and as a reference tool—requires an understanding of how a compliant eCTD is structured and the process changes needed for successful implementation. Presently, many organizations are submitting improper eCTDs, either because they failed to learn the intricacies of what constitutes a validated eCTD, or have not adopted the necessary process changes. FDA can reject improper eCTD submissions for technical reasons, meaning that the clock for review does not begin, thereby delaying approval. When this happens, the submitter is notified to make corrections.

Antoinette Azevedo, president and CEO of **e-SubmissionsSolutions.com**, says a key barrier to eCTD acceptance is companies' lack of budget, followed closely by non-compliant content preparation. The three most common errors she sees are: invalid XML content; inability to navigate the PDF files that constitute the narrative part of the submission; and datasets submitted in invalid format. These problems are slowing the adoption of eCTD and hampering regulatory review. But

Azevedo attributes these errors to fixable rookie mistakes and to the fact that the system is still flooded with paper. "The challenge is that because of the paper, there is a lot of legacy content, and there are some people who think that the best way to get to eCTD format is through scanning," she says.

Much of what Azevedo says is confirmed by FDA representatives, who peg scanned documents as particularly frustrating to reviewers, as they are not searchable and are often illegible. Other common errors are a missing or inaccurate table of contents, inoperable hyperlinks, and missing files (see "Common Errors in eCTDs").

According to *Specifications for eCTD Validation Criteria*, an FDA publication, incoming eCTDs receive a programmatic validation step that checks the backbone against more than 100 rules. Errors are rated as high, medium, low, or ignore, which determines the level of reviewability. A "high" error would be missing files, the lack of the study report, protocol, and methods validation, without which the review cannot proceed. Similarly, if the

submission does not have a US-regional.xml file, which identifies application type, it is impossible to process it. In July 2008, EMEA established its own validation criteria using 44 rules, rating violations as serious, medium, or low.

While submitters have the opportunity to make corrections, a better route is to avoid these problems altogether. Improving performance requires process changes. Booth-Genthe says that Wyeth identified its first challenge as getting documents from the authoring department in an acceptable format. This required ensuring the documents used the correct font, had appropriate hyperlinks and margins, and had the granularity that eCTD requires.

Granularity refers to the level of the smallest file unit or building block within the eCTD. As a highly defined structure, the eCTD includes a hierarchy of files and folders as laid out in the table of contents. The idea of granularity is to avoid combining several documents

that violate the eCTD structure, confuse reviewers, and hinder life cycle management down the road.

"There was some resistance in the chemistry, manufacturing, and control (CMC) organization to break down the documents to that granular level," says Booth-Genthe. "There was a very long-standing way of writing those sections, and we had to change that. We had to change the templates, the guidances, and the SOPs. It was a gradual process."

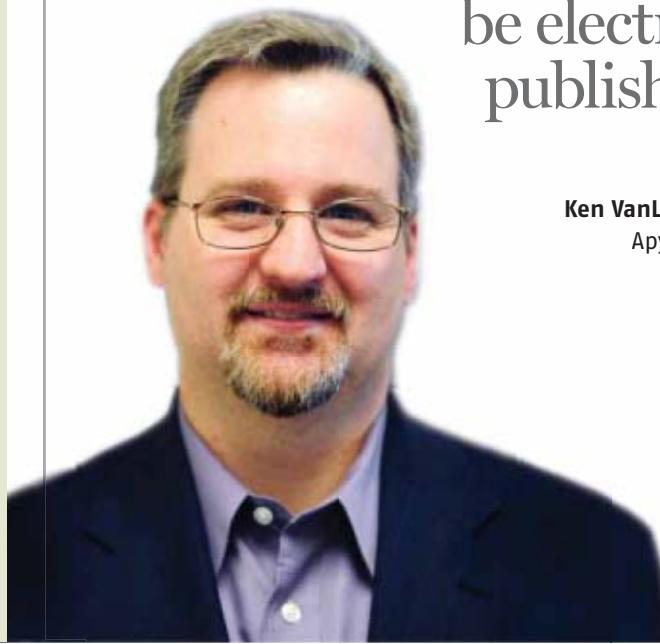
As companies begin implementing the necessary changes, there is optimism that the quality of submissions is starting to improve. Errors still abound, but there is growing awareness among trial sponsors that electronic submissions cannot be reviewed if they are of poor quality.

Virginia Ventura, regulatory information specialist in CDER's Office of Business Process Support, has noticed improving submission quality because more industry representatives are sending email questions to the ESub support

Common Errors in eCTDs Reported by FDA

- Documents scanned instead of using PDF, and are often illegible
- No table of contents or an inaccurate one lacking bookmarks
- Inoperable hyperlinks
- Missing files or empty folders
- Incorrect or repeated use of application number
- No US-regional.xml file
- Level of granularity is not in keeping with the table of contents
- Use of spaces and characters not permitted by the Electronic Submissions Gateway
- Sending submissions to CDER instead of CBER, and vice versa

"In approximately three years... Apyx anticipates that 65% to 70% of company business will be electronic publishing."



Ken VanLuvanee,
Apyx

Feature

staff (email: esub@fda.hhs.gov), submitting sample eCTDs, keeping abreast of updated specifications, and reviewing recent FDA eCTD presentations. "Documents are being placed in the correct or more appropriate locations within the eCTD, eCTD metadata is accurate most of the time, and leaf titles are more descriptive in the eCTD tree as compared to previous submissions," Ventura remarks. Sponsors who obtain clarification from the EsuB Support Team or ensure eCTD format is included in the pre-NDA or pre-IND meetings have a higher incidence of making successful submissions.

Get Ready for RPS

As sponsors cope with the best way to expand eCTD submissions, they will soon be confronting what many see as the next big trend—implementation of the Regulated Product Submission (RPS) standard. The objective of RPS is to define a single message structure that can be used globally for all regulated products, including biopharmaceuticals, medical devices, veterinary medicines, and food.

RPS is an HL7 standard that offers many of the same advantages of the eCTD plus additional strengths. It offers two-way communication between the submitter and the regulatory agency, whereas eCTD is a one-way communication tool. In addition, with RPS, previously submit-

ted documents will not have to be resubmitted in marketing applications, as it will be possible to apply previously submitted documents to the marketing application. This will spare submitters from having to retest hyperlinks and bookmarks because the agency will already have those documents. Lifecycle management will also improve with RPS as it will be possible to change granularity and move documents, steps that are not possible with eCTD.

Jason Rock, CTO of **GlobalSubmit** explains how RPS will impact eCTD. "The next major version of eCTD is going to transition to the RPS standard. The same content submitted in eCTD, such as the PDF documents and the SAS datasets, will be submitted in RPS, but the inner workings of XML will be completely changed."

According to Rock, by September 2012, FDA must have RPS implemented because it has Prescription Drug User Fee Act (PDUFA) commitments for two-way communication, and is planning on using RPS to meet those needs. The next milestone is January 2010, the date by which standards are to be in Draft Standard for Trial Use (DSTU) form, or ready for implementation on a trial basis. FDA is slated to develop an RPS Implementation Guide by the end of 2010, and expects to be able to accept RPS Release 2 submissions in spring 2011.

Looking Ahead

Although RPS is looming, eCTD adoption should continue to rise, especially as more regulatory agencies build an infrastructure to accept them and begin encouraging or mandating their use. Anticipated growth is driven by the fact that the majority of submissions are still made in paper. According to the EU eCTD Implementation Survey, by June 2008, a mere 1% of submissions used eCTD format. In the U.S., as of the end of 2007, only 11% of the 168,000 submissions to FDA used some form of electronic method.

The ability of companies to make process changes that will enable increased eCTD submissions, either through building in-house expertise or through outsourcing, is being hampered by economic concerns and the need to contain costs. Toban Zolman, director of consulting services at Image Solutions, says, "Companies are trying to solve those challenges without laying out a lot of capital investment, and by taking some novel approaches to the technology." Larger companies are more likely to implement their own solutions in house, and may outsource portions based on region or submission type. Smaller players are taking a blended approach, looking at Software as a Service (SaaS) and hosted options. Helping drive down costs are the movement toward harmonization and fewer regional nuances, and an increase in the number of vendors providing eCTD services. These factors will eventually create more of a commoditization of the eCTD.

As eCTD becomes more integral to the global submission process, companies stand to gain the largest benefit as they start building in eCTD requirements from the initiation of the Phase I studies. This will result in higher quality electronic submissions from the beginning, cutting out paper, the costs associated with it, and allowing for faster regulatory review. •

Further Reading:

Specifications for eCTD Validation Criteria, March 10, 2008, www.fda.gov/cder/regulatory/ersr/validation_1.0.pdf

Robinson-Kuiperi C. Ensuring a Successful Submission. Office of Business Process Support, Center for Drug Evaluation and Research, FDA, November 5, 2008, www.fda.gov/cder/regulatory/ersr/SanDiegoTutorialPart2.pdf

**"There was some resistance...
to break down the documents
to that granular
level."**

Terri Booth-Genthe,
Wyeth

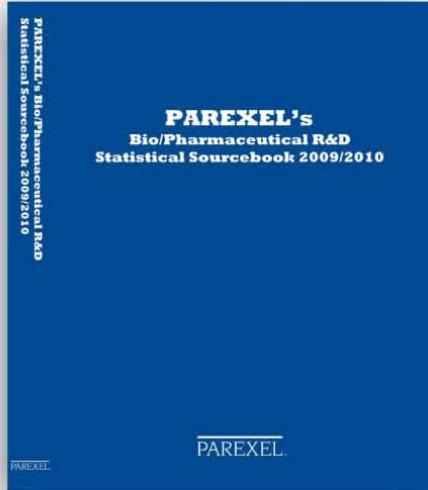




Data...Analysis...Trends...Industry Insights

PAREXEL's Bio/Pharmaceutical R&D Statistical Sourcebook 2009/2010

Hot Off The Press...New Edition!



"PAREXEL's Bio/Pharmaceutical R&D Statistical Sourcebook is one of the most important sources for accurate industry and regulatory data, and for some tables, it is the only source."

—**Bert Spilker, President
Bert Spilker & Associates, LLC**

The PAREXEL's Bio/Pharmaceutical R&D Statistical Sourcebook 2009/2010 is the leading resource for statistics, trends and proprietary market intelligence and analysis on the biopharmaceutical industry. Supported by thousands of graphs, illustrations, and analysis, the sourcebook provides the latest intelligence on every aspect of biopharmaceutical development – from product discovery to R&D performance and productivity to time-to-market trends.

With key analysis and contributions from leading consultancies and experts, the sourcebook provides real-world data and analysis, including:

- New proprietary analysis on US clinical trial starts, segmented by therapeutic category, as well as overall active clinical trials
- Emerging data on worldwide and company-specific R&D pipelines and product launch trends
- An all-new and comprehensive analysis of clinical research offshoring revealing which pharma companies are now locating their new clinical trials
- New analysis on emerging trends in pharma and biotech licensing deals and other partnerships critical to industry's efforts
- Drug approval statistics compiled from FDA, EMEA, and other regulatory agencies
- New global R&D spending trends and other international R&D data from key markets
- And much more!

"This is an incredibly useful and reliable resource for anyone connected to the pharmaceutical and biotech industries: marketers, financial planners, investors, or researchers."

—**Judy Blaine, Librarian
ArQule, Inc., Business Information Alert**

Considered a 'must-have' resource for the drug development industry, it is an invaluable resource for executives and managers working within or with the pharma and biotech industries and puts real-world data sets at your fingertips for presentations, reports, business development efforts, strategic meetings, and critical decision-making analyses.

The 2009/2010 edition will also be offered in electronic format.

To purchase the PAREXEL's Bio/Pharmaceutical R&D Statistical Sourcebook 2009/2010,

Call 1-800-856-2556, Email customer.service@barnettinternational.com, or Visit the Publications section on www.barnettinternational.com

Barnett International: A division of Cambridge Healthtech Institute, www.barnettinternational.com.
250 First Avenue • Suite 300 • Needham, MA 02494 USA • Phone: (800) 856-2556



www.barnettinternational.com

Bio-IT World

Signature Series

ADOBE

Deploying Adobe Technology to Automate Electronic Submissions

Life sciences organizations are rapidly moving toward eCTD as the electronic submission format worldwide. As of January 1, 2008, sponsors providing electronic submissions to the FDA for NDAs, INDs, ANDAs, BLAs, Annual Reports, and Drug Master File submissions are required to submit in eCTD format. Waivers are available for sponsors that are unable to submit in eCTD format and paper is still accepted. As of July 1, 2008, eCTD is the preferred electronic format for submitting dossiers in the European Centralized Procedure, and many European Authorities are well ahead of the targeted 2009 date for accepting eCTDs.

LiveCycle ES provides Life Sciences companies with technology that can be implemented incrementally to automate the creation of regulatory submission components and achieve optimal process efficiency over time, allowing for a modular approach to achieving full eCTD capability. All companies, from large R&D companies to small biotech or contract research organizations, can realize significant savings across product development efforts. Since the processes are repetitive and consistent, once process efficiencies are implemented, they can be leveraged across all development efforts, resulting in ongoing savings and faster time-to-file across all therapeutic areas.

LiveCycle ES is an integrated server solution that blends data capture, information assurance, document output, process management, and content services to enable the delivery of enterprise applications for life sciences. LiveCycle ES solution components include:

- Adobe LiveCycle PDF Generator ES: Enables the generation of agency-compliant PDF files
- Adobe LiveCycle Forms ES: Enables the creation and deployment of interactive forms such as review and approval
- Adobe LiveCycle Digital Signatures ES: Enables the automated creation and validation of digital signatures, including SAFE-BioPharma digital signatures, in PDF files
- Adobe LiveCycle Rights Management ES: Enables document-level version and access control of Adobe PDF, Microsoft Word, and Excel files so confidential information is restricted to intended recipients
- Adobe LiveCycle Reader® Extensions ES:

ROI

Delays in finalizing submission components lead to delays in launching your product, resulting in lost revenue. Manual processes make it costly, and labor-intensive to adhere to internal procedures and regulatory mandates such as CFR Part 11.

Enables the use of free Adobe Reader software
 • Adobe LiveCycle Process Management ES: Streamlines manual, time-consuming processes such as review and approval
 • Adobe Live Cycle ES Connectors for ECM: Enables service operations for creating and updating content items, locating

items by their attributes, and retrieving documents and their attributes

Adobe Reader is one of the main user interfaces for LiveCycle ES applications. When supported by components of LiveCycle ES, Adobe Reader users can complete forms, add comments and annotations, and digitally sign documents. Since Reader is freely available, LiveCycle ES offer solutions for secure workflows that move both inside and outside the firewall.

The component creation process is typically well defined and repetitive within each organization lending itself to automation across all therapeutic areas. In many companies, the creation of a submission component depends on manual processes to support the information in the report. Improving the management and automation of this process can deliver efficiencies to the overall submission process.

LiveCycle ES provides the tools to smoothly integrate existing systems and improve processes. Examples:

- LiveCycle Process Management ES
 - Automate the retrieval of appropriate SOPs.
 - Automate the retrieval of appropriate templates.
 - Automate the routing of documents to and from primary authors and reviewers or contributors.
 - Help ensure the capture of audit data surrounding the authoring process.
- LiveCycle PDF Generator ES
 - Automate all regulatory PDF file generation from the server to promote the creation of consistent PDF files in accordance with predetermined settings.
- LiveCycle ES Connectors for ECM
 - Automate the updating of content in the submission management system.
 - Interface with other content repositories

to help ensure that the required documentation and metadata are entered into or retrieved from the repository(ies).

In many organizations the routing of components for review and approval (R&A) is a time-consuming, error-prone, and labor-intensive process. Manually routing components for R&A requires that one or more individuals assume responsibility for managing and monitoring the process which can increase the overall timeline and introduce errors and/or delays.

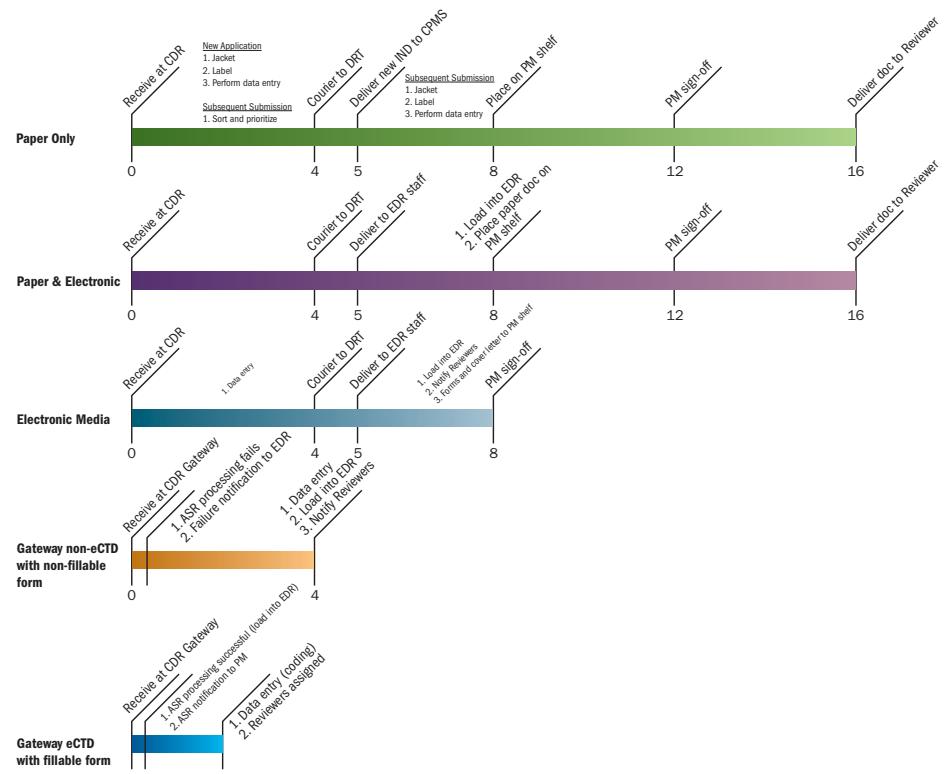
Intelligent documents helped reduce forms processing time from five minutes to one minute. Digital workflow helps ensure reliable document delivery shortening time for processes by several days.

Merck KGaA deploys Adobe technology to optimize clinical trials for new pharmaceuticals.

The following LiveCycle ES components can be applied to automate the R&A process and enable a more secure and compliant online collaboration environment:

- LiveCycle Forms ES
 - Automate the generation of customized R&A forms to be routed to all reviewers.
- LiveCycle Reader Extensions ES
 - Enable users of Adobe Reader to save documents, add comments, and digitally sign documents (including SAFE-BioPharma digital signatures).
- LiveCycle Process Management ES
 - Identify reviewers.
 - Automate reviewer notification that R&A process has been initiated.
 - Monitor R&A status.
 - Collect audit information about R&A processes.
- Live Cycle ES Connectors for ECM
 - Automate the updating of content in the submission management system.

Document Processing Timelines



FDA Statistics presented at DIA EDM Meeting, February, 2009, indicate that document processing time is reduced from 16 days to 2 days when eCTD submissions are sent via the FDA Gateway

– Interface with other repositories to help ensure that the required documentation and metadata are entered into or retrieved from the repository(ies).

As a trusted enterprise partner, LiveCycle ES helps you streamline both paper-based and electronic elements of the submissions process—including document creation, collaboration, assembly, review, and approval. At the same time, it enables you to secure the movement of information to and from contributors, across business processes, and between sponsors and regulatory agencies worldwide. As a result, your organization can speed its delivery of more accurate and more secure information for compliant electronic submissions.

"We chose to go with a single solution for all research areas, so there would be one searchable archive of PDF files and a common set of tools for everyone. We knew we couldn't get the return we wanted if we didn't adopt a fully electronic solution. Adobe LiveCycle solutions were integral to our success."

Kay Gross, Former Senior PKI Specialist, Information Security and Solutions, Proctor & Gamble

CROs Heed the Crisis Call

The CRO market is growing despite economic climate.

BY REBECCA PALMER

The CRO market has grown dramatically over the past decade, with world-wide revenues estimated to be \$17.8 billion and an annual growth rate of around 15% since 2006. Yet, there are hurdles ahead as pharmaceutical partners endure mergers, lackluster pipelines, generic competition, and rising health care costs. Costs must decrease, and emerging scientific models must be reevaluated for biological relevance, cost efficiency, and international policy compliance.

The economy could be cause for worry, but John Lewis of the [Association of Clinical Research Organizations](#) disagrees. "Initially, we saw a slowdown in research into early 2009, but eventually, that work must be done. I think we will see pent up demand. The focus will be on core competencies—the pharmaceutical companies will do what they do best, the drug discovery and the marketing, while they leave the middle piece, the development, to us. So signs of increased outsourcing look good."

Indeed, Watertown, Mass.-based [Apredica](#) is seeing "unprecedented demand" according to its president, Katya Tsaioun. As pharmaceutical companies tighten their belts, CROs can expect additional work. Indeed, [Covance's](#) Tom Privette, VP strategic partnering, sees "increased urgency", plus a need to do "more with less" and the requirement to "access more expertise."

But the economic crisis has hurt CROs in other ways. Tsaioun remarks that many of Apredica's customers have collapsed due to their failure to obtain funding. "A specter of looming insolvency [has been cast] over much of the industry," she says. "If investment funding doesn't start flowing back into the industry soon, by the end of the year things could turn for the worse." [PPD Inc.'s](#) chief operating officer William Sharbaugh agrees, saying, "tighter credit markets have caused some biotech companies to cancel or delay projects to conserve spending. The overall number of biotech companies is decreasing because of acquisitions or bankruptcies."

The Obama Administration's stimulus bill could be just what the CROs ordered. Tsaioun comments: "In the past month [since the stimulus bill], we've had more requests for consulting, service proposals,

and support letters for grant applications than we had in the entirety of 2008." Covance's Privette thinks that more understanding about the stimulus bill is needed, but the company that won "Best Contract Research Organization of 2008" for "best-in-class services to support vaccine development" at the 2008 World Vaccine Congress, is heartened by the promise of enhanced funding for preventative medicine.

Turning to E-Tools

To help serve their customers' needs of efficiency while maintaining a global reach, CROs are turning to electronic tools to share data with sites throughout the world. Covance, which carries out 25% of the world's clinical trials, offers clients the opportunity to build their own or buy pre-made electronic data capture (EDC) solutions to fit any kind of study, company strategy, or budget. These solutions allow data to be collected and aggregated for an entire project, regardless of geographic location or tests performed. Covance's clients can access information within 24 hours, from more than 100 countries. PPD has used EDC solutions in trials for over six years, equaling more than 200 clinical trials, according to Sharbaugh. PPD prefers Oracle's RDC Onsite 4.5.3, which

allows for faster data entry through a web-based interface.

The Internet, too, plays a large part. Long-distance communication made possible by the Internet permits Apredica to speak to their clients throughout the study process in areas such as India, Australia, and South America. As Tsaioun says, "The common perception is that outsourcing is a one-way street, but that's not so. When customers become globally convenient to work with, customers focus on what company is the best provider for them."

The economy is not the only challenge facing CROs. Governments are applying their own pressures on the market. Policy makers around the world are bending to public pressure to reduce animal testing. The European Union, for instance, has a mandate that no consumer products (e.g. shampoo, cosmetics) tested on animals can be made or marketed in Europe after 2012. Apredica's Tsaioun says "our industry is in the massive process of replacing expensive and imprecise *in vivo* testing with more accurate and less costly tools... since the mid-1990s we have been making dramatic, rapid advances in *in vitro* testing."

Yet pharmaceuticals that treat life-altering and sometimes fatal diseases are a far cry from cosmetics. Despite public aversion to animal testing, drug makers must ensure their products are safe. Any new testing model must give strong indications of efficacy and toxicity prior to commencing clinical trials. More work clearly needs to be done in order to ensure that non-animal based models are as predictive, or more predictive, than current animal-based models. Tsaioun agrees and says, "Clinical candidate selection needs to happen not only based on efficacy but also on strongly supported *in vitro* tests that are predictive of human clinical outcomes."

Despite the stiff challenges faced today, CROs are finding strategic solutions to bridge the gap between pharmaceutical companies and the public.



There's **ePRO**.

And then there's
ePRO from ERT.



Now there's one-stop to fully engage your patients throughout the trial lifecycle. **ERT's proven ePRO Solutions** drive subject recruitment, retention, compliance, data capture, and even safety monitoring. No more running on empty when it comes to optimal study conduct – **ERT's ePRO** will remove traditional roadblocks and accelerate study success, in real time.

www.ert.com/ePRO

► To test drive **ERT's ePRO**, call +1 866 538 2808 or email epro@ert.com today.

IT/Workflow

Information You Can Bank On

The UK Biobank hopes to manage and mine a quarter of a billion data points.

BY VICKI GLASER

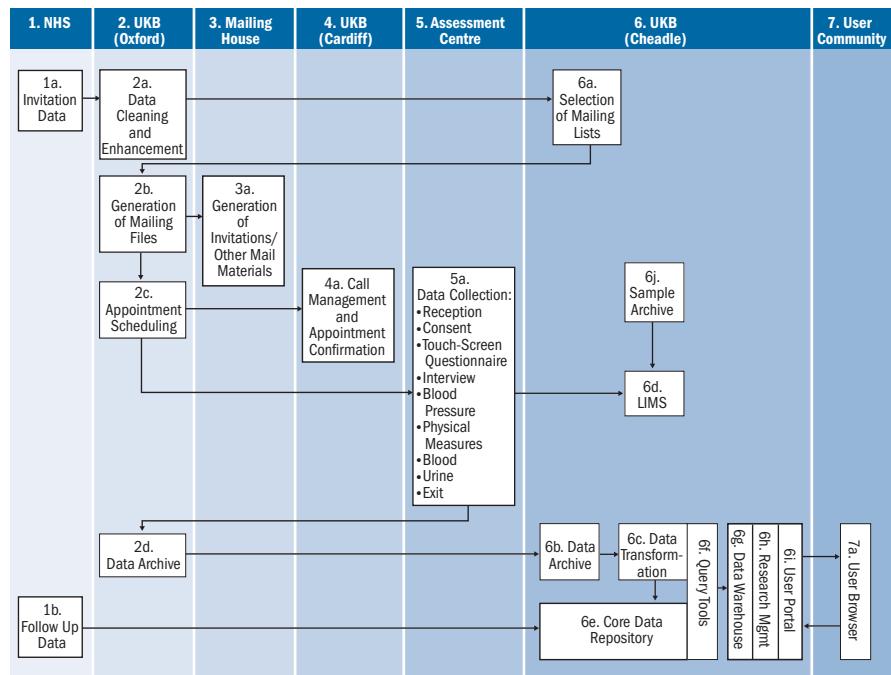
Andy Harris, Information Services Director of **UK Biobank**, was entrusted with a daunting task when he joined the huge data repository project, which was first conceptualized a decade ago: Design an IT system with the capability to manage more than a quarter of a billion data points, track over 20 million blood samples, and, perhaps most challenging, analyze and mine all that data without knowing how the information might be used 5, 10, or 20 years down the road.

And one more thing: all this information is highly personal and must be collected, stored, and retrieved in a secure, limited-access system that preserves and protects the anonymity of the registrants.

Harris viewed this challenge as the Rubik's Cube of system design—get all the pieces of the puzzle in place and the reward would be a boundless resource for future clinical trials. A repository of personal demographic information, linked to medical history, physiological parameters, and blood samples, which together could form the basis for a better understanding of disease predisposition and the development of screening tests, targeted diagnostics, more accurate prognostic measures, and new therapeutic strategies.

From its inception, the driving force behind the creation of the UK Biobank was “a strategic need to establish large blood-based prospective epidemiological studies with prolonged and detailed follow-up of cause-specific morbidity and mortality,” states the Website. This vast database has a set of goal of 500,000 registrants, with approximately 500 data points for each individual—250 million data points to be collected, stored, retrieved, and analyzed (see “Jewels in the Biobanking Crown,” *Bio•IT World*, May 2007).

A massive challenge is to ensure that at every step, the pivotal considerations are privacy, security, and ethics. These



The UK Biobank's systems architecture is designed for security.

overriding principles guide every programming detail, keystroke, and decision about access to a cache of highly sensitive information. Ensuring an unalterable link between an individual and his or her evolving data file, while not revealing the person's identity, is a challenge. Guaranteeing the integrity and security of the data as it is collected at temporary Assessment Centers throughout the country and transferred via the Internet between geographically dispersed processing and archival sites is a headache. And determining the best way to store this data to optimize its value as a resource for future medical research, when it is unclear how the information will ultimately be used, is an IT team's nightmare.

“We have to be open-minded at every stage, while making sure we are meeting all of the ethical and security requirements,” says Harris, who has led the software development and IT management team for more than four years. The

team consists of six people who built the IT infrastructure from scratch, with design decisions and system modifications implemented as the concept evolved. Many of the day-to-day data management tasks are outsourced to contract service providers.

UK Biobank chose to build patient records using the **Oracle** application Healthcare Transaction Base (HTB), designed to be HL7-compliant. The IT team developed Oracle- and Java-based software to convert the data upstream of HTB. Once data are in the repository, they are Oracle-dependent. In addition to HTB, Biobank uses two main commercial products: Thermo Nautilus LIMS and **Microsoft** Exchange BackOffice function, which facilitates communication among the staff.

Protect the Innocent

Harris's team deliberately did not design the system architecture as an interactive

environment, intending to limit access to data stored in the core systems. The group developed an IT system specifically to manage assessment center visits to register new participants. Each center typically has 30 to 40 touch-screen desktop workstations linked through a local area network (LAN) to an on-site server. Only the center's server and the administrator's unit have Internet access.

When new registrants arrive clutching their invitation letters, they are given a USB stick containing an encrypted key that translates the name, address, and identifier printed on their letter to their assessment specific identifier. "Linkage of name to identifier is a crucial part of the infrastructure," says Harris.

A nurse inserts the USB stick into a workstation, enters the identifier, and completes a detailed consent process that includes a digital signature. Participants then answer some 300 questions on a touch-screen module before an interactive session with a nurse, who enters information on medical history, medications, and physical measurements (blood pressure, height, weight, bone density, lung function, etc.). After obtaining six vacutainers of blood and one of urine, the nurse gives participants a copy of their signed consent form and the basic results of their tests.

"We now have sensitive data in the assessment center computers," says Harris, which are encrypted locally using the Blowfish algorithm. Every 20 minutes, new information is transmitted through encrypted tunnels over the Internet to the UK Biobank central server in Oxford. With each successful transmission, the operator receives a confirmation flag and the data packet is automatically cleared from the Center's workstation.

Dedicated software also guides the processing of the blood samples. The vacutainers are bar-coded with color-coded caps corresponding to how each will be processed and stored. Once a blood sample is registered, the software initiates a timed process, with a visual progress bar and periodic reminders to guide operators through the processing steps. The LIMS tracks the samples, records how and where they are stored, and links the associated test results. The samples are shipped overnight to the Biobank center

Banking 101

- Completed registry will contain information and blood samples for 500,000 Britons between the ages of 40 and 69 years at time of entry.
- Infrastructure started in 2003-04; data collection began in April 2007; more than 250,000 people already enrolled.
- Some 10% individuals who receive an invitation letter respond.
- Each Assessment Center operates for 6-12 months before moving to another city (rolling data collection).
- Biobank registers about 3,500 new participants/week; about 14 staff can process more than 100 people/day; baseline assessment takes about 90 minutes.
- UK Biobank maintains operations in Oxford, Manchester, and Cardiff.

in Manchester.

Every night, all the data transferred that day to the central server are re-encrypted and loaded onto a secure web server at the Biobank center. At this point, "The data has finally arrived on the edges of the repository," says Harris. The information is also archived on tape at two sites, including Oxford. In the future, follow-up data from primary care records, hospital visits, and test results maintained by the UK's National Health Service (NHS) will be appended to the Biobank's core data repository and linked to pre-existing data. The Biobank also hopes to gather data on death certification and cancer registration.

Online Banking

Although the UK Biobank may include some free text entries to establish the exact nature of a medical decision, it will primarily rely on capturing coded information. For every item entered, the IT team had to develop terminologies that map a code back to the words entered.

For example there is a terminology set for diagnoses borrowed from the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes.

Merging free text with codes presents dilemmas. For example, one participant might enter "Holland" as his birthplace, another "The Netherlands." Different entries cannot share a single code, but there is reluctance to alter the source data because, as Harris points out, an accurate audit trail "is critical to clinical trials." The Biobank solution is to create a new "version 2" record, which translates "Holland" to "The Netherlands" and assigns the appropriate code.

The data stored in the central repository reside in XML format, because of a desire to be compatible with the NHS switch to Health Level 7 (HL7), a framework for the integration and exchange of electronic health information. This standard provides a common language for communicating clinical data. "UK Biobank chose to use it rather than develop our own standards," says Harris. Although he describes HL7 as being "verbose and extensive," it offers the advantage of being system independent.

When a participant's consent materials and baseline information arrive on the periphery of the repository, it is checked for proper encryption and completion. Next, the system creates an internal message in HL7 to register the person in the repository. This serves as a "person record" and creates a portal for entry of clinical data, though no information (except for the consent) is stored there yet.

Harris's group has been developing an extraction tool that can retrieve data and format it into an output that would make sense to an epidemiologist trying to link participant parameters to disease susceptibility, diagnoses, and outcomes.

"We are seeing remarkable results," says Harris. "We can turn around 125 million data points in about 5 minutes. It's not real-time, but it's heading in that direction." The retrieval system is still in development, Harris adds, and "the ethical and security issues have to be wrapped around it," to ensure that the data are linked to the correct person and that only approved people can access the data. •

IT/Workflow

Isilon's Data Storage Odyssey

CEO Sujal Patel and the quest to “remove storage as an impediment to progress.”

BY KEVIN DAVIES

During his five years working at Real Networks, the pioneering streaming media Internet company, Sujal Patel began to appreciate the sheer size and growth of the data his customers were trying to store. Their frustrations sharpened his ideas on how to solve the storage scalability problem. In 2001, he founded **Isilon Systems** in Seattle, with the goal of building a storage architecture from scratch. Patel realized this file-based explosion was much broader than digital content, reaching all sorts of data—oil and gas, media, design, and bioinformatics.

In 2007, Patel took over as CEO after the company went public. Since then, Isilon has reacted to what Patel calls the “avalanche of life sciences data,” attracting marquee clients both in academia (Broad Institute, UCLA) and industry (Complete Genomics). Nowhere is that growth more evident than in the data-intensive next-generation sequencing space. The proportion of life science customers grew from 2% to 12% in 2008, while total revenue almost doubled in two years, pushing above \$100 million. The company’s client base will soon top 1,000. “We have an incredible cash balance (\$80 million) to innovate during this time,” says Patel. “There’s still a lot of work to be done to take storage out as an impediment to progress.”

On a recent visit to *Bio•IT World* headquarters, Patel unveiled a new range of storage products. Isilon is purely a storage company, says Patel, but in the broadest sense. “Storage for us is about data protection, data management, long-term data analysis, anything having to do with data purely at the storage level.” When asked why launch the products now, Patel replies simply: “We’re done building them!”

There are two big problems facing life sciences organizations, says Patel. “One is a capture problem—how do I get the data from the sequencer, get it in and start

to work with it? Then, even larger, is the huge repository of the genome sequences that you want to keep online for a long time, refer to them and analyze them. That’s the larger opportunity.” For example, Patel says the problem facing Complete Genomics is how to generate and store 30 petabytes (PB) of human genome data online for a year? “It’s a huge task,” he admits.

Isilon’s traction in the biomedical community can be traced to several factors. “Isilon has a scalable architecture that allows you to add both performance and capacity incrementally and thus scale to where your application demands you to be,” says Noemi Greyzdorf, research manager with IDC. Traditional NAS [network-attached storage] approaches limit users to the performance of a single node. Clustered file systems are typically associated with high-performance computing environments “but the advances in applications in the biomedical world and other verticals, have increased demand for that architecture in more commercial applications.” Moreover, “Isilon has done a lot of work on making it user friendly so a typical storage/NAS administrator can maintain the environment as any other NAS.”

Designing Storage

Patel realized in founding Isilon that regardless of the data type—sequence data, video clips, chip designs—storing huge repositories of fast-growing files required a much more scalable architecture than what existed at the time. “We leveraged principles of clustered computing to create from scratch a whole new architecture for storage, built around the needs of these huge stores of file-based data. When you have a huge store of file-based data, you have to get it onto the storage quickly, grow and scale effectively, utilize



Sujal Patel

all hardware effectively and simplify that whole architecture. We did that and it intersected exactly with next-generation sequencing requirements.”

While “very sophisticated” clients such as the Broad Institute (see “A Broad View,” *Bio•IT World*, Apr 2008) and Complete Genomics are among the more prestigious clients, there are many more customers with impressive needs of their own—a few sequencers and 1.5 PB storage. “For anyone who knows how to manage storage and move data around, the data protection strategy, the simplicity of our solutions is something that customers in this space really like,” says Patel.

Providing solutions for the life sciences market, Isilon draws on experiences in other tiers. “In the semiconductor space, it’s getting to the point where chip designs are getting so complex that the needs out of the storage system have grown by an order of magnitude. That sort of growth parallels the same sort of thing we’re seeing in life sciences and next-gen sequencing, the same sort of thing we saw a couple of years ago in media.”

Ideally, Patel says, it would be nice to have “a big Z drive in the sky, you’d put my genome, your genome, and so on. But if you look at NetApp and EMC, the two

(CONTINUED ON PAGE 50)



The Path to Drug Development, Regulations, and Industry Trends

NEW! PAREXEL's Bio/Pharmaceutical R&D Statistical Sourcebook 2009/2010

Considered 'a must-have resource' for the drug development industry. This statistical sourcebook is filled with pharma/biotech R&D trends, data, market intelligence, articles, and graphs providing fresh insights into the developments reshaping pharma R&D and the drug development industry. An invaluable resource for executives and managers working on presentations, reports, business development efforts, and critical decision making analyses. Also available in electronic format.



Coming Soon! The U.S. Drug Approval Trends and Yearbook 2009/2010

A one-of-a-kind compendium filled with hundreds of key trends and metrics used by professionals to gain industry insights, new benchmarks, and analysis to help plan their own R&D projects, improve company performance, and assess various drug approval options and strategies.

New Drug Development: A Regulatory Overview 2008

Addresses the most cutting-edge developments redefining how new drugs are developed and regulated today. The book is considered the "go-to" resource for regulatory, clinical, project management, training, and other drug development disciplines navigating the FDA's drug development approval processes.

Expediting Drug and Biologics Development: A Strategic Approach

Using a unique "reverse-engineering" approach, dozens of leading experts with extensive experience in all disciplines of drug and biologic development show how careful planning and a sharp focus on the end-goals can be used to expedite the most complex product development program.

Biologics Development: A Regulatory Overview

The only text book on biologics development and regulation in the post-CBER/CDER Consolidation Era! Offering an expansive examination of the FDA's regulation of biologic products, from preclinical testing to post-marketing regulatory requirements, and from user fees to electronic submissions.

Regulatory Reporter Newsletter

A monthly newsletter providing regulatory information for the pharmaceutical, medical, and allied industries. Annual subscriptions available.

Barnett International: A division of Cambridge Healthtech Institute, www.barnettinternational.com
250 First Avenue • Suite 300 • Needham, MA 02494 USA • Phone: (800) 856-2556



Purchase your copy(ies) today for your colleagues, research staff members, and professionals involved in the management and conduct of clinical research.

To order, call 1-800-856-2556 or visit the "Publications" section on www.barnettinternational.com

- Emerging pharma/biotech R&D trends
- R&D performance and productivity
- Phases of clinical development
- Informed consent & IRB review
- Integrated approach to data management
- Nonclinical drug testing
- The Biological License Application (BLA)
- Priority drug review destinations
- CDER and the NDA review process
- Pediatric studies initiative
- Post-approval changes to marketed drugs
- Phase 4 commitments

www.barnettinternational.com

IT/Workflow

Isilon Storage

(CONTINUED FROM PAGE 48)

leaders in the NAS space, the maximum volume of size they can build is 16 TB.... I can't even fit a genome onto one volume."

That necessitates an "an extremely manual process of scaling up hundreds of these separate volumes—that's an incredible amount of complexity and leads to poor utilization of the storage systems. It means you're buying twice as much as you need. And you have a ton of staff micro-managing this daily."

Isilon's storage typically requires less people to manage the load balancing, scalability, and hardware upgrades, says IDC's Grezdorf. "The clustered architecture allows you to automatically load balance users and data across the nodes in the cluster. It also allows you to scale simply by adding more nodes instead of doing a forklift upgrade to the next model," she says. Moreover, upgrading hardware is also seamless as the data are simply swapped to other nodes in the cluster. The system automatically rebalances the data across the new node.

According to VP marketing Ram Appalaraju, the Broad Institute formerly had three people managing the institute's NetApp infrastructure full time. Now it's about 1/3 of a person. At last year's supercomputing conference, Broad Institute CIO Jill Mesirov said the institute had been using eight NetApp filers and 54 Sun Thumper filers before it moved 92% of its workflow onto Isilon. "Now they have 1.5 PB on the same cluster," says Appalaraju. "Because we're driving utilization up to 80%, our 1.5 PB is equivalent to 2.25 PB of our competitors. You get much more of your money's worth."

New Series

Isilon has added three new products to its existing X-Series nodes, providing a tiered set of storage solutions. To create and capture data, it has introduced the S-Series—the IQ 5400S—which allows customers to build clusters. This will appeal to "those who need high performance but don't typically have a lot of data that need to be processed at the same time," says Grezdorf.

Patel says the S-Series "is going to be

incredibly important for high-speed data analysis, such as coming off next-gen sequencers." That said, Patel thinks the S-Series unit is probably faster than most of the apps currently being targeted in life sciences. But, "If we get to the point where the sequencers output grows another half order of magnitude than the current generation of machines," then all bets are off. "I bet you we're going to run into at least a few life science customers where the performance needs get up high enough we've got to go this route," he says.

For archiving, it released the NL (nearline) storage IQ 36NL. The NL-Series disk-based deep archiving solution economically scales up to 3.45 PB within a single file system, at a unit price of about \$2/GB, with 80% storage utilization. And rounding out the X-Series is the new top-of-the-line IQ 36000. The X-Series suits those with lesser performance needs but more capacity. It allows Isilon to scale up to 2 or 3 PB in a single cluster. It is a 4-U box (two processors per box) with 36 TB per node, that can be clustered together to create extremely large file systems with high performance. It provides up to 30 GB/second performance.

"From the data management perspective, we can uniquely tailor a particular product while still bringing in overall benefits of scalability," says Appalaraju. Says IDC's Grezdorf: "It is designing to the need instead of trying to position the same product for all types of needs."

2010

Patel signs off with a flourish, pointing to the remarkable specification enhancements over the past few years that put Moore's Law to shame. "If you look at 2010, our vision is to be able to build 10-, 15-, 20-PB clusters with performance that's virtually unlimited, so we can take away the challenge of storage being an inhibitor to applications," he says.

"Our key focus is around data management, data protection, and ensuring our customers have a well rounded offering—you plug it in, turn it on and it goes. From a hardware perspective, it means you'll see us do things like integrate SSD drives instead of hard disc drives, and improve our hardware technologies. These are all things that are in the works." •



5400 S-Series (create and capture)

- Enterprise Class 2U Isilon IQ Node
- 12 450-GB 15K RPM SAS
- 5.4 TB storage per node
- Dual, quad-core 2.33 GHz CPU
- 16 GB SDRAM
- 4 x 1 GbE interfaces
- DDR InfiniBand cluster Interconnect



IQ 36000 X-Series Node (Process)

- Enterprise Class 4U Isilon Platform Node
- 36 1-TB 7200 RPM SATA 3Gb/s drives
- 36 TB capacity per node
- Dual Quad-Core Intel 2.33 GHz CPU
- 8 GB SDRAM
- 4 x 1 GbE Front-End Interface
- DDR InfiniBand for intra-cluster communication
- 7-node minimum cluster configuration (252 TB)



IQ 36NL (Archive)

- Enterprise Class 4U Isilon Node
- 36 1-TB 7200 RPM SATA 3Gb/s drives
- 36 TB capacity per node
- Single, Quad-Core Intel 2.33 GHz CPU
- 4 GB SDRAM
- 1 x 1 GbE Front-End Interface
- DDR InfiniBand for intra-cluster communication
- 7 node minimum cluster configuration (252TB)

Clinical Research

Social Networking Sites Embrace Clinical Trials

Patient recruitment is online and interactive.

BY DEB BORFITZ & ALLISON PROFFITT

More and more online social networking sites are assembling groups around particular disease types as a way of simulating interest in clinical trials and the data they're collecting on members. It's a recycled idea, one that was abandoned by Veritas Medicine. But earlier approaches failed in part because "the recruitment and retention process relied upon other players outside the control of the matchmaker companies," says Kenneth Getz, chairman of the [Center for Information and Study on Clinical Research Participation](#).

Getz says patients were also disappointed to find a "highly limited listing" of clinical trial options and were "uncomfortable divulging personal information to for-profit companies." Moreover, low consumer awareness of clinical trials made it difficult to reach sufficient numbers of study volunteers. "Ultimately, the per-patient cost of online matchmaking couldn't justify the investment."

New social networking sites hold more promise, but can they successfully attract study volunteers, given limited awareness of trials or familiarity in social networking? As Getz puts it, "Is the value in their reach or in the data they are collecting on community members?"

Inspired

[Inspire](#) was created four years ago to accelerate recruitment via social networking. "You can't create a community just about clinical trials," says CEO Brian Loew. Inspire can deliver "high quality patients," screened based on a protocol's inclusion and exclusion criteria, because it has a pre-existing relationship with members who have volunteered to be trial participants.

The first step was to build online com-

munities around a particular disease, which Inspire has done in partnership with dozens of non-profit health associations. Trial recruitment began last year and Inspire already has four contracts in place, two with a major pharma for trials in osteoporosis, arthritis, and lung cancer. Loew says the early indications are good, adding that for one of the cancer studies, Inspire already has a community with ten times the number of members who meet the initial inclusion criteria.

To the delight of trial sponsors, Inspire offers a "quality assurance" service whereby trial participants are surveyed about their satisfaction with the investigative site. Responses get reported, in aggregate, to companies. The fact that Inspire owns and operates the communities assuages industry's liability concerns regarding conversations about off-label use of their products and adverse events, says Loew.

Patients Like You

[PatientsLikeMe](#) focuses on five chronic-illness categories, including amyotrophic lateral sclerosis (ALS), and takes a different recruitment approach that is also winning support. This no-advertising venue efficiently targets "those who are both eligible for and interested in information about clinical trials," says president and co-founder Ben Heywood.

The U.S. Medical Outsourcing Group at [Novartis](#) partnered with PatientsLikeMe for a pilot multiple sclerosis (MS) clinical trial awareness program launched last May to enhance recruitment efforts for the FTY720 FREEDOMS II trial. The pilot consisted of a monthly email sent to MS patients, encouraging them to visit the institutional review board- (IRB-) approved website ([MSClinicalTrials.com](#)) for prescreening to determine eligibility. A Novartis spokesperson called the experience "very positive."

The screenshot shows the Inspire website homepage. At the top, there is a header with the Inspire logo and a tagline "together we're better". Below the header, a banner features a quote: "You have helped me more than I can express. In talking with everyone else I have found hope and courage". The main navigation bar includes a "Search" button and a "Find it" link. The page is divided into three main sections: 1. **Search**, which includes a search bar and a link to "Find health information and discussion in 502,569 member posts. Learn more..."; 2. **Join**, which shows icons for various groups like "Osteoporosis", "Arthritis", "Lung Cancer", "FSR", "Heart", "Stroke", "Diabetes", and "A"; and 3. **Connect**, which shows icons for "Groups", "Friends", "Events", and "Photos". Below these sections, there is a call to action: "Connect with 105,573 members and make friends who share your health interests, learn about treatments, find support and more. Learn more...".

www.Inspire.com

Heywood is particularly intrigued by industry's idea "to use our platform to create a micro-community for patients participating in clinical trials." The idea would be to capture real-world data between site visits, providing useful information on safety, dosage, and perceived efficacy prior to regulatory submission. "This could ultimately lead to more accurate labeling and safer treatments for patients," says Heywood.

[PatientsLikeMe](#) offers its sponsor partners a clinical trial awareness package, which might include aggregated data about patients in the community they're targeting. "We are also in discussions with pharma about collaborating to supplement trial data they collect with patient outcome information." [PatientsLikeMe](#) can also tell its partners about how their approved treatments are being used in the real world.

Interestingly, [PatientsLikeMe](#) initiated the first "patient-driven, real-world natural experiment" to learn if lithium arrests the progression of ALS (as suggested in a report last year). The experiment used validated outcome scales to quantify the patient experience on lithium while monitoring side effects and blood levels, allowing the company to amass what Heywood calls "the largest population of ALS patients that has ever been assembled for a study of this kind from which to analyze data."

Click it Forward

[Click It Forward](#), started by [Acurian](#), uses established social networks to spread the word about trials. Facebook and MySpace

Clinical Research

The screenshot shows the homepage of PatientsLikeMe. At the top, there's a navigation bar with links for "Patients", "Treatment", "Symptoms", and "Research". Below the navigation is a search bar with placeholder text "Search this site" and a "Search" button. A "Help" and "Click" link are also present. The main content area features a "Share Your Experience" section with a "Get how easy it is..." link, a "Find Patients Like You" section with a "Get started" link, and a "Learn From Others" section with a "Discover the power of many..." link. There's a "Join Now (It's free!)" button and a note that "Already a member? Log in". Below these sections are several small thumbnail images of users. On the left, there's a sidebar titled "Our Current Communities" listing various medical conditions like "Neurological Conditions", "Hematological Conditions", "Parkinson's Disease", "ALS (Amyotrophic Lateral Sclerosis / MND)", "PMS", "P.D.", "Primary Lateral Sclerosis", "MSA", "Progressive Nonfluent Aphasia", "PTSD (Post Traumatic Stress Disorder)", "CBD (Cannabidiol)", "MSA (Multiple System Atrophy)", and "Duchenne Neuromuscular Disease". On the right, there's a "Highlights" section featuring a "Fibromyalgia Community" with a "Get involved" link, a "Lyme & ALS Study" with a "Get involved" link, and a "Clinical Trial Awareness" section with a "Get involved" link. A note at the bottom left says, "If we have a community for your condition, join now! Don't see your condition? Request it here." A quote from Angie Wells, Co-founder, HGA ALS Division, is at the bottom right.

www.PatientsLikeMe.com

users sign up for the application and as a user's friends sign up, Acurian will donate money to one of 20 medical causes chosen by the user.

"It's a two-fold mission," says Scott Conner, Acurian's VP of marketing, of the database that is currently 50 million strong. The first goal of Click It Forward is to "grow more acute indications... and build channels to people who have diseases or know someone who has a disease." Conner's second goal for the program is more expansive. "We want to increase public awareness of clinical trials."

The free application is available through www.clickitforward.org. Users register and install the application on their Facebook or MySpace account. They then invite their friends and contacts to add the application as well. The application also integrates with Google Earth so that users can see a dynamic, visual map of their own Click It Forward network. Users gain points based on how many friends subsequently install Click It Forward. The money that Acurian donates to the user's chosen cause is based on ten tiers of accumulated points.

The contribution angle is meant to, "encourage people to make it viral," Conner explains. Acurian has committed to giving up to \$50,000 per year to causes chosen by Click It Forward members. Members choose from over 20 common diseases or indications including breast cancer, fibromyalgia, migraines, pain, depression, and diabetes. Acurian will

The screenshot shows the homepage of Click It Forward. At the top, there's a navigation bar with links for "GET INVOLVED", "FAQs", and "ABOUT ACURIAN". Below the navigation is a banner with the text "SUPPORT YOUR CAUSE. SPREAD THE WORD." and a large photo of a smiling family. The main content area features a "The First Click" section with a quote from Kate, a mother of two, about her family's health challenges. Below this are three bullet points: "My brother, with high blood pressure", "My son's soccer coach, with rheumatoid arthritis", and "My pastor, just diagnosed with multiple sclerosis". To the right, there's a "That's why I joined Click it Forward" section with a "GET STARTED" button. On the left, there's a "Highlights" section with a "Fibromyalgia Community" link, a "Lyme & ALS Study" link, and a "Clinical Trial Awareness" link. On the right, there's a sidebar with "Get involved", "Frequently Asked Questions", and "About Acurian".

www.clickitforward.org

identify "appropriate, nationally recognized non-profits" at the end of the year to receive the money allocated to each cause. Members also have the option to choose secondary causes about which they are interested in receiving more information.

Acurian hopes to target two different demographics with Facebook and MySpace. "Facebook is an older audience," Conner says, usually presenting with more chronic indications. MySpace, on the other hand, caters to much younger users. "They are the future of clinical trials," Conner says, and he hopes that Click It Forward will work to, "condition them to the importance of clinical trials."

Acurian manages their database of 50 million individuals who have expressed interest in clinical trials, listed their health problems, and given detailed contact information. The company works with trial sponsors in a "soup to nuts" fashion to run recruitment campaigns, screen patients, and track the back end of the process.

An Expanding Field

Online marketer [Alliance Health](#) launched its first social networking site, DiabeticConnect, last June to promote clinical trials. So far, 5% of its 17,000 registered members have opted to be notified of relevant trial opportunities, says CEO Stead Burwell. Alliance Health also drives traffic to trial screeners via email, newsletters, and online advertisements. The trial matching service is three times more effective

than other types of interactive channels in terms of successful screens, he adds. This year, Alliance Health plans to launch a new networking site every four weeks, including ones for sleep disorders, obesity, chronic pain, and heart disease.

[Inclinix](#) has learned firsthand that social networking sites can be highly effective and affordable recruitment venues, says chief marketing officer Chris Sleat. "Many sponsors fear bloggers will turn against their trial, but they'll blog positive about [a company] as long as [it is] open and honest." Among the sites favored by Inclinix are Twitter, Reddit, Digg, Craig's List, YouTube, Facebook, and MySpace, plus multiple European sites.

For more than a year, patient recruitment firm [MediciGlobal](#) also has been successfully using social networking sites like Facebook and MySpace, according to president and CEO Liz Moench.

Meanwhile, more than 1,000 clinical researchers and health care professionals are collaborating on the Vision Tree Optimal Care (VTOC) Network. This site allows physicians to "interact, refer, and discuss current issues and clinical trials" at no charge, says Adam Hawkins, director of technology. It integrates with Twitter and YouTube to share blog posts, files, and links between members. "We are working with partners who have a large patient population to allow for outreach and recruitment into clinical trials through the VTOC Network. You will see some of these in coming months." •

LIFE SCIENCE REPORTS YOU CAN RELY ON

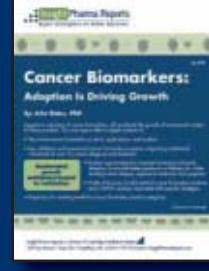
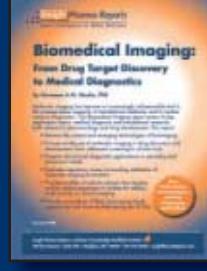
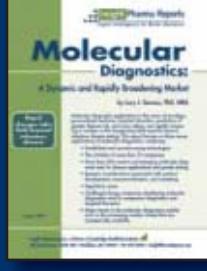
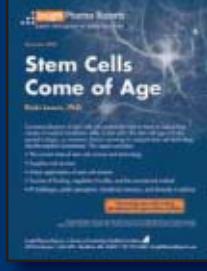
Timely • Insightful • Concise



Advance your knowledge with **Insight Pharma Reports'** collection of published reports that provide timely, insightful, and concise analysis on specific topics for your drug development needs. The reports help professionals meet the need to stay abreast of the latest advances in pharmaceutical R&D, their potential applications and business impacts, and their current and future position in the marketplace.

To order your copy(ies) of these new titles or other titles, call 781-972-5444, email rlaraia@healthtech.com, or visit InsightPharmaReports.com.

An Essential Collection of Timely & Concise Reports Analyzing the Technologies, Markets, and Strategic Issues Driving R&D Productivity



Order your copy(ies) today!

Visit InsightPharmaReports.com for other available reports.

Corporate License and Corporate Subscriptions are available. Contact David Cunningham at 781-972-5472 or Cunningham@healthtech.com for special program packages.

"CHI's Insight Pharma Reports are cost-effective and a reliable source of information about the markets we serve. The reports allow us to keep on top of developments in this rapidly moving field."

-Donald N. Halbert, Ph.D., EVP of R&D, Iconix Bioscience

Published by



Insight Pharma Reports, a division of Cambridge Healthtech Institute
250 First Ave, Suite 300, Needham, MA 02494
T: 781-972-5400 or Toll-free in the U.S. 888-999-6288 • F: 781-972-5425

Clinical Research [GUEST COMMENTARY]

The Art of Texting in Clinical Trials

Texting offers instant contact and feedback.

BY TIM DAVIS

Cell phone technology is a ubiquitous means of personal and professional communication and is now developing into a workable solution for the clinical trials arena. With more than 4 billion cell phone users worldwide, the technology is helping the pharmaceutical and health care sectors maintain direct communication with their clinical trial subjects.

Many large pharma companies are implementing cell phone technology as part of the clinical trial communications process—from subject recruitment through to collecting trial data. Text message alerts are a quick and effective way of communicating—with all cell phones having built-in text messaging capability. Using cell phone technology, companies can increase recruitment and retention rates by interacting directly with trial subjects.

Here, I examine the current strategies of subject recruitment, retention, and compliance during clinical trials; how problems can be overcome with the use of cell phone technology; and what the future holds for the technology within the pharmaceutical industry.

Bridging the Communication Gap

Effective recruitment is essential to a successful clinical trial. Conventional forms of recruitment including physician referral have a very low success rate. Advertising in the media, universities, hospitals, and clinics is a popular recruiting method, while some organizations prefer to write letters to potential participants inviting them to take part in studies. Traditional methods can be useful during the recruitment process; however they require the subject to take the first step.

Working with communication providers, sponsors can interact directly with subjects in a personalized and cost-effective way. Cell phone technology integrated into the clinical trial communication process at an early stage opens up direct com-

munication channels. Existing advertising campaigns can be enhanced by printing a short code on advertisements, encouraging subjects to text details relating to a specific trial. Potential participants can then be directed to call centers or self-enrol using their cell phone. Recruitment results are delivered instantaneously compared with traditional methods where results might not be seen for weeks or months. Cell phone technology can offer recruitment response rates of over 30% compared with typically 10% for more traditional methods.

Non-compliance costs the pharmaceutical industry billions of dollars each year. Significant costs can be incurred by sending out reminder letters or making telephone calls to subjects to ensure they comply with the requirements of the study. This is another area that can be tackled using cell phone technology to distribute automatic reminders directly to the subject. Messages can be tailored to the individual to include information such as dosage, frequency, and how to take the medication safely. Delivery reports or interactive messages can be employed to ensure that messages are received.

Ongoing support and motivation is often needed to support participants throughout a clinical trial. Sponsors use various methods to keep subjects involved in a study including providing detailed paper-based information on the medication and side effects at the outset of a study. In order to keep participants well-informed of the aims of the trial and to remind them of the benefits of remaining in the study, short messages can be sent to motivate and enhance confidence

in the trial.

Cell phone technology offers a powerful method of direct communication and can be used to overcome the most common challenges in clinical studies. Subject reporting via text message can help to improve data quality and effectively determine the efficacy of a new drug.

Cell phone communication strategies implemented alongside existing procedures or as a stand-alone method, benefit from significantly improved overall data quality and optimised use of resources.

Of course, security and confidentiality are important considerations and all methods of data collection and evaluations must be developed in accordance

with strict regulatory requirements—FDA 21 CFR 11 and HIPAA in the U.S. and EU Directive 2002/58/EC in Europe. Subject consent should be obtained and an opt-out option should always be made available. All information must be encrypted and stored on a secure database in order to ensure full compliance with the EU data privacy act.

Analysts predict that text messaging technology will be superseded by new methods of cell phone communication, such as 3G, which enables more detailed and personalized data to be collected. As new cell phone technologies emerge, they will be developed and adapted for use in the pharmaceutical industry. Vendors in the sector will increasingly need to invest R&D activities in mobile technology to provide the most effective and compliant solutions. •

Cell phone technology integrated into clinical trials at an early stage opens up direct communication channels.

Tim Davis is co-founder and CEO of **Exco InTouch**. He can be reached at tim.davis@excointouch.com.

Coming Soon!

Biomarker BRIDGE

An interactive collaborative network community for pharmaceutical, clinical and IT professionals who are interested in gaining and sharing knowledge about technologies, research, and regulatory issues around life science.



Join and become a member of the Biomarker Bridge - enrollment opens soon!

First 200 charter members will receive a complimentary subscription to Biomarkers in Medicine

- Create a personal profile
- Browse the membership library
- Complimentary access to CHI Biomarker Events and Webinars*
- Search continually updated Speaker Abstracts White Paper and Poster library
- Network with your peers through Forums, Blogs and Polls
- View continually updated Biomarker News
- Receive on-line access to life sciences publications

* Membership level determines access

Membership Options:

There are several different types of membership access to the Biomarker Bridge. Members to the Biomarker Bridge are granted access to a password-protected web site filled with information, tools and files specifically for Biomarker Bridge members.

Interact with your peers and engage in member-driven research, presentations, webinars and more.

The *BEST* decision you will make in 2009

www.biomarkerbridge.com

If you have questions, comments or suggestions for the Biomarker Bridge, email:
biomarkerbridge@healthtech.com

Biomedical Research Initiative Discussion Groups for Excellence

Educational Opportunities

This section provides a variety of educational events in the life sciences industry that will help you with your business and professional needs. To list an educational event (print & online), contact Sales_chmg@chimediagroup.com. To preview a more in-depth listing of educational offerings, please visit www.bio-itworld.com

Featured Events

DIA 45th Annual Meeting, June 21–25, 2009, San Diego, CA Better Medicines: Improving Safety with Every Step

No matter what your area of expertise, or interest area, this meeting has been designed specifically with you in mind. Network with more than 8,500 colleagues from several geographic locations, professional interest area and level of responsibility; Listen to more than 1,100 speakers, including representatives from FDA, EMEA, and other global regulatory agencies; Choose from more than 350 sessions, 34 tutorials, and 3 mega-tracks; and much more. Visit www.diahomes.org

CHI Events

For more information on these conferences and other CHI events planned, please visit www.healthtech.com

RNA Interference Summit
June 8–10, 2009 • San Francisco, CA
Genomic Tools & Technologies Summit
June 8–10, 2009 • San Francisco, CA
The Business of Genomics
June 8–9, 2009 • San Francisco, CA



World Pharmaceutical Congress
June 9–11, 2009 • Philadelphia, PA
Collaborative Innovation in Biomedicine
June 22–23, 2009 • Washington, DC
Trends in Cancer Diagnostics
August 10–11, 2009 • Washington, DC
Clinical Adoption of Next Generation Diagnostics
August 11–12, 2009 • Washington, DC

Novel Vaccines: Design & Development
August 17–18, 2009 • Providence, RI

The Bioprocessing Summit
August 24–27, 2009 • Cambridge, MA

HCA East: Drug Discovery & Development Series
September 21–23, 2009 • Boston, MA

Exploring Next Generation Sequencing
September 21–23, 2009 • Providence, RI

ADAPT 2009: Accelerating Development and Advancing Personalize Therapy
September 22–25, 2009 • Washington, DC



Bio-IT World Conf & Expo Europe
October 6–8, 2009 • Hannover, Germany
PEGS Europe
October 6–8, 2009 • Hannover, Germany
Accelerating Proof of Concept
October 13–14, 2009 • Boston, MA
Immunogenicity Assessment and Clinical Relevance
October 26–28, 2009 • Philadelphia, PA

Barnett Educational Services

Visit www.barnettinternational.com for detailed information on Barnett live seminars, interactive web seminars, on-site training programs, customized eLearning development services, and publications.

Live Barnett Seminars

Patient Recruitment and Retention: Successful Planning and Management

June 8–9, 2009 • Philadelphia, PA

Managing and Conducting Global Clinical Trials

June 10–11, 2009 • Philadelphia, PA

Working with CROs

June 23–24, 2009 • Boston, MA

Pharmacovigilance in Europe: An Introduction for US Personnel

June 29–30, 2009 • Philadelphia, PA

Medical Device Postmarketing Vigilance Reporting

August 27, 2009 • Chicago, IL

Managing and Conducting Global Clinical Trials

August 27–28, 2009 • Chicago, IL

Patient Registry Programs
October 21–22, 2008 • San Francisco, CA

Interactive Web Seminars

Meeting International Safety Reporting Requirements: Addressing the Challenge for Both Large and Small Organizations
June 10, 2009

The Pharmaceutical and Medical Device Industries Today: Comparing and Contrasting
June 16, 2009

Biomarkers of Drug Efficacy and Safety
June 19, 2009

Electronic Medical Records and Source Document Verification
June 24, 2009

Quality Systems: A Controlled Approach to GCP Compliance
June 29, 2009

Monitoring Phase I Clinical Trials
July 1, 2009

**Reference Manuals...
Industry Compendiums...
Regulatory Resources...
Clinical Guides**

Barnett International, a division of CHI, is a well-known publisher of reference guides and publications focusing on the drug development, regulatory, and clinical research industry.



Detailed information on these valuable books can be found on www.barnettinternational.com

 BARNETT EDUCATIONAL SERVICES

White Papers, Webcasts, and Podcasts

Keep abreast of current industry trends and developments. Browse our extensive list of free Life Science white papers, webcasts, and podcasts on www.bio-itworld.com. Interested in learning about developing a multi-media solution to generate leads? Please email Sales_chmg@chimedigroup.com

White Papers

Reducing Bottlenecks in GMP Laboratories

Sponsored by Waters

Significant time is spent conducting quality control (QC) testing and documenting QC results during the pharmaceutical manufacturing process. GMP regulations require maintaining documentation to ensure strict compliance with established SOP's. Completing paper documents and ensuring their authenticity creates a bottleneck in the QC laboratory. Waters® NuGenesis® SDMS Intelligent Procedure Manager, an electronic SOP workflow & documentation system, addresses the predominately manual activities required to perform an analytical method.



Coherent IT Strategy for Fast-Moving R&D

Sponsored by Tessella

Putting value on an IT strategy is easier with an end-to-end view, from target to market, of the cause and effect of good and bad information and decisions. The white paper covers how: a) biopharma companies can build effective IT for R&D strategies; b) relevant systems can be deployed around a particular set of goals; and c) case studies have been conducted resulting in a better business return.



Applications of LIMS for Enhanced Productivity in Translational Discovery Research

Sponsored by Ocumim

Translational research can improve the effectiveness of drug research and target discovery pursuits. However, managing the massive amounts of data has posed enormous challenges for -omics and biorepository labs. This white paper discusses the latest developments in translational



research and describes how an adaptive data infrastructure and a flexible Lab Information Management Solution (LIMS) framework can be pivotal for successful Translational Research.

A Rapidly Deployable Data Management System for Clinical Research

Sponsored by LabKey Software

The Statistical Center for HIV/AIDS Research and Prevention (SCHARP) statisticians and internal IT staff created an Integrated Data Management System for Clinical Research working closely with open-source LabKey Server systems. The system enables collaborators to more rapidly analyze, validate, and share datasets from labs and clinics around the world, as they are collected.



Translational Research Briefing On Report

Sponsored by GenoLogics

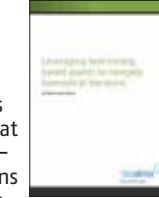
The varied collection of Bio-IT World articles and insights assembled in this Briefing ON examines the impact of integrating clinical data back into the research and development pipeline; and using information gathered from physical samples, databases, and clinical trials to benefit the design and performance of future research.



Leveraging Text-Mining Based Search to Navigate Biomedical Literature

Sponsored by Bioalma

With the amount of research data available today estimated at 18 million documents across more than 20,000 medical journals, it is more important than ever that students, researchers, professionals, doctors, and librarians have easy access to the latest research information available. The white paper covers the pros and cons of various ways in which to search biomedical information; and how novoseek – a new biomedical search engine – leverages text mining to offer an easy to use, and efficient search system for biomedical literature and grant information.



Webcasts

Adobe Secure Web Conferencing for Healthcare & Life Sciences

Sponsored by Adobe

Web conferencing is now established as one of the most essential tools for business communications. From collaboration around new drug development and clinical trials to

marketing new healthcare products and research, Web conferencing can save time and money while increasing productivity. The webcast provides a comprehensive overview of the features and benefits of Adobe Acrobat Connect Pro for secure web conferencing in the healthcare and life sciences industry.



Increase the Pace of R&D Innovation with Google

Sponsored by Google

Life sciences organizations are increasingly looking to IT and better tools to improve researcher productivity and shorten the product development process.

As the leader in search technology, Google has developed a universal search for enterprises, which provides the ability to search all enterprise content—including Intranets, file shares, databases, real-time business data, and content management systems—through one simple search box.



Podcasts

What Will Drive Tomorrow's Health Business Innovation: Technology, Standards, or New Business Models?

Sponsored by Medidata

Executives from Medidata and SAS discuss the current state of health business innovation, particularly around providing solutions for clinical trials and drug development and address the following questions:

- Does the health industry need more technology to drive business innovation?
- How does the current state of industry standards impact the ability of technology to drive business innovation?
- Is the lack of integration, inability to enter clean data or data privacy issues posing barriers to fully exploiting today's technology?



Next Generation Sequencing Data Management

Co-Sponsored by Genomics and Illumina

Targeting genomics labs with increasing volumes of data and complexity of experiments, this end-to-end solution helps manage their entire process from receipt of samples to integrated data for analysis. Learn how the fully integrated Genome Analyzer-Geneus solution can help maximize your next generation sequencing investment.



The Russell Transcript



Microsoft's Amalga Life Sciences

JOHN RUSSELL

One of the biggest commercial announcements at the just concluded Bio-IT World Conference & Expo was Microsoft's introduction of Amalga Life Sciences. This is the next step in Microsoft's ambitious plan to create a single information highway and data aggregation/modeling platform—Amalga—that spans basic research, clinical trials, health care delivery, and consumer health information needs.

"We think of this as an information supply chain where information can flow in the right way, at the right time, and in the right context. Think of five broad categories of activities: collect, validate, store, analyze, and share. Amalga Life Sciences can navigate through this data and connect even dissimilar forms of data, such as biochemical data, EMR, pathway data, gene ontologies, expression data, and so on," says Jim Karkanas, senior director of applied research, Microsoft Health Solutions Group.

Amalga's roots stretch back to 2006 when Microsoft purchased Azyxxi, an enterprise platform developed for hospitals by a group of physicians to address the issue that few of the disparate systems in hospital networks share data in ways that allow physicians or administrators to see everything at once. To some extent, the release of Health Vault—Microsoft's Personal Health Record product—is the tip of Microsoft's Amalga consumer vision of the supply chain.

A guiding principle in the development of Amalga (Azyxxi) is that memory (hardware) is inexpensive and the best way to deal with the myriad data types generated in health care settings is to capture all data and metadata at the point of generation; store it in a consistent, computable form; and provide easy interfaces and modeling capabilities.

Amalga LS pushes the platform into basic research and clinical trials, with the goal of managing and analyzing heterogeneous research data. The Amalga LS "hybrid store" manages

simple and complex data—experimental protocols and higher-level knowledge representations—thereby increasing productivity. Among Microsoft's early users are the [Fred Hutchinson Cancer Research Center](#) in Seattle and the [David H. Murdock Research Institute](#) in Kannapolis, North Carolina.

"Our researchers face an overwhelming challenge to collect, analyze, interpret, and share complex data from a wide range of diseases and experiments," says Nobelist Lee Hartwell, president and director of "the Hutch," who hopes Amalga LS will help drive understanding of the data "and ultimately help us meet our vision of enabling personalized medicine."

Managing Connections

Another early Amalga program participant is the [University of Trento Centre for Computational and Systems Biology](#) (CoSBi), which is 50 percent owned by Microsoft (see "CoSBi Models," *Bio-IT World*, Mar 2009). CoSBi develops languages and mechanisms of modeling and simulations in medical science and biology. One can imagine CoSBi-developed technology finding its way into Amalga.

Amalga LS features a "Concept Browser" for the UMLS (unified medical language system) ontology, which enables users to input and connect multiple concepts via their ontological relationships.

Users may choose to connect via all possible connections, to select the depth of connections, or to display only the shortest path. The output is displayed as a visual network of concepts and relationships. Once the data are accessible, subject matter experts can then apply models and theories to data, identify the best solutions to complex problems, and predict new relationships.

As an example, Karkanas offers a pharma modeling genomic data collected in clinical trials. "Those modeling tools only can look at the data in the very narrow sense of what they were designed to look at and can't really correlate that information with the EMR data that might be available for these patients. That's a connection we would be able to make with our platform," says Karkanas.

Microsoft insists it won't compete with application providers, building an EDC system for example. Yet some functionality (e.g. modeling) in the platform does intrude into the application space. Amalga LS will include an API to allow other vendors to build plugins for the platform.

Amalga is hardly Microsoft Office for the health sciences. Still, Microsoft is putting together an interesting strategy flexible enough to accommodate ever-advancing tools as on- and off-ramps to its health information highway and repository, while building in fundamental application-like capabilities (e.g. modeling) into the roadway and database architecture.



The Microsoft booth at the Bio-IT World Expo introduced Amalga and Health Vault.

MARK GARBRECK

For Knowledge Mining and Data Management ...



... you need **GeneGo**

Systems Biology and Pathway Analysis

GeneGo's high quality manually curated database and suite of software solutions cover all the major diseases: Metabolic Diseases, CNS, Cancers, Cardiac, Immunology and Inflammation – plus Stem Cells

MetaCore™ and **MetaDrug™** offer an easy to use interface that provides high quality results fast! Data mining is available via EZ Search where you can type in a gene, protein, compound or microRNA and within seconds find which canonical pathway maps or prebuilt networks they are involved in. Alternatively, you can use MetaSearch to build a Boolean query to find pertinent information from GeneGo's manually curated underlying database, MetaBase™. With MetaSearch you can find all the kinases in breast cancer that are not in prostate cancer or find compounds with a certain value range for IC50's.

GeneGo software applications are easy to use and include free on-line training to get you up to speed quickly.

Get high quality answers to your data mining questions in seconds using **MetaCore!**

Test it yourself today with a FREE trial – trials@genego.com or genego.com/trials



500 Renaissance Drive, Suite 106
St. Joseph, Michigan 49085 USA
(888) 592-3124, (858) 756-7796
+44 (0) 7525 064993 United Kingdom
www.genego.com
sales@genego.com

© Copyright 2009 GeneGo Inc, all rights reserved worldwide. GeneGo and MetaCore are trademarks of GeneGo Inc.



Take the shortest path to the **RIGHT PATH.**



Symyx Isentris

The Freedom to Explore.

Out here, being able to make decisions quickly and confidently changes more than your perspective. It gets you to the right path faster. The same holds true in the lab. With Symyx Isentris® software, delays caused by technology and overwhelming amounts of data are a thing of the past. Instead you discover more, repeat less, and get it right the first time.

Symyx Isentris lets you search, browse, analyze, organize, and report on corporate, commercial, and local database information following a self-service model. You can share data, collaborate with colleagues, and gain valuable insights. The result is less time sorting data and more time exploring science.

*To learn more, visit
www.symyx.com/isentris*

*To attend a free webinar, go to
www.symyx.com/webinars*



© 2009 Symyx is a registered trademark of Symyx Technologies, Inc. All rights reserved.