



Web 2.0 and the Future of Competitive Intelligence

Text Analytics, Portfolio Analysis, and the Real-Time Value of Digitised Content

Stephen A. Leicht

Chief Operating Officer

www.collexis.com

Key Messages...

- For the tech buyers - If you can dream it, 'we' can do it
- For the tech sellers – GUI, GUI, GUI
- For the big content players – buy, partner, or perish

A moment on definitions – Web 2.0 and 3.0...

What is the semantic wave?

A tidal wave of four Inter

How is Web 3.0 different from previous stages of internet evolution?

Knowledge computing driven and solves problems of scale

What semantic technologies will power Web 3.0?

Digital tools that represent and reason about meanings, theories, and know-how separately from documents, data, and program code.



The semantic wave embraces four stages of internet growth. The first stage, *Web 1.0*, was about connecting information and getting on the net. *Web 2.0* is about connecting people — putting the “I” in user interface, and the “we” into Web of social participation. The next stage, *Web 3.0*, is starting now. It is about representing meanings connecting knowledge, and putting these to work in ways that make our experience of internet more relevant, useful, and enjoyable. *Web 4.0* will come later. It is about connecting intelligences in a ubiquitous Web where both people and things reason and communicate together.

The basic shift occurring in Web 3.0 is from information-centric to knowledge-centric patterns of computing. Web 3.0 will enable people and machines to connect, evolve, share, and use knowledge on an unprecedented scale and in new ways that make our experience of the internet better.

Web growth continues to accelerate. Dimensions of net expansion include communications bandwidth, numbers of people connected, numbers and kinds of devices that are IP-aware, numbers of systems and applications, quantities of information, and types of media. As the internet expands, needs world-wide are outstripping the capacities and capabilities of current information and com-

The key notion of semantic technology is to represent meanings and knowledge (e.g., knowledge of something, knowledge about something, and knowledge how to do something, etc.) separately from content or behavior artifacts, in a digital form that both people and machines can access and interpret. As a platform, Web 3.0 will embrace all semantic technologies and open standards that can be applied on top of the current Web. It is not restricted just to current Semantic Web standards.

Web 3.0 will encompass a broad range of knowledge representation and reasoning capabilities including pattern detection, deep linguistics, ontology and model based inferencing, analogy and reasoning with uncertainties, conflicts, causality,

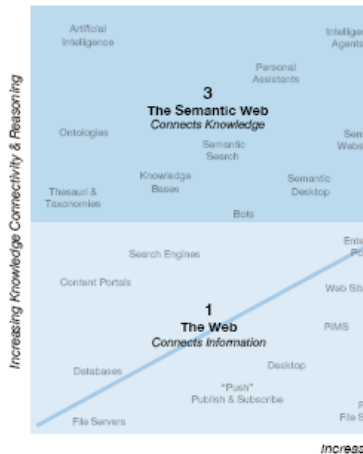
and values. The figure below depicts a spectrum of progressively more capable forms of knowledge representation that spans tag collections (or folksonomies); to dictionaries, taxonomies and thesauri; to schemas and conceptual models; to ontologies and theory-based logics, to axiologies (value-based reasoning), and entirely new uses barely tapped. Reasoning requires knowledge representation. We choose more powerful forms of representation to enable more powerful kinds of reasoning and problem solving. The integration of social Web and semantic technologies in Web 3.0 allows new synergy that lowers the cost of data and knowledge creation, and raises the computational value of gathering it.

Executive Summary 5

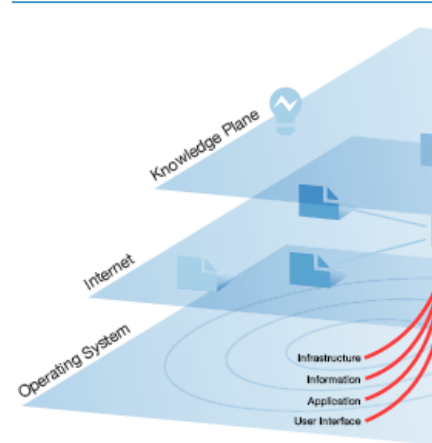
Proj
Se
Inc
Mu

EXECL
Februa

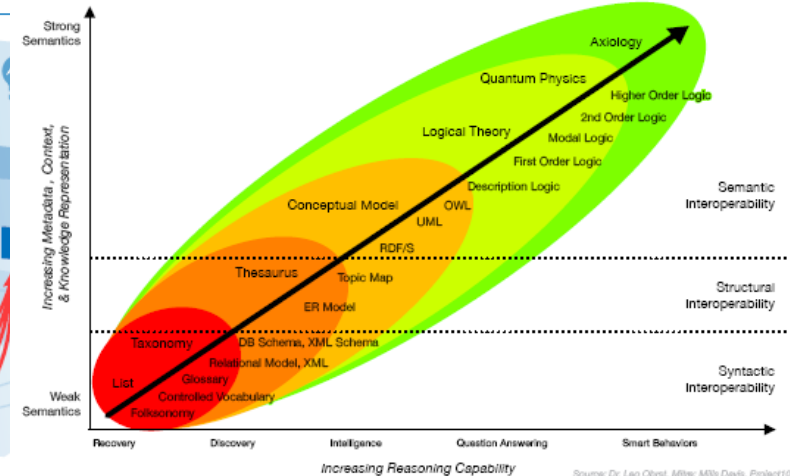
Mills D
www.p
202-66



Below:
Web 3.0 — The Internet Grows a Knowledge Plane



Below:
From Searching to Knowing — Spectrum of Knowledge Representation and Reasoning Capabilities



Source: Dr. Leo Obrst, Mills; Mills Davis, Project10X

2007, 2008 Copyright MILLS+DAVIS. All rights reserved

A Little History

Data and Search...(the story of a terabyte)



Good news – we aren't going away...

Outsell estimates the overall information industry to be worth \$387 billion in 2007, having grown 6.6% from its 2006 value of \$363 billion. Outsell tracks 12 segments within the information industry, as detailed in Table 1 and Figure 1.

Table 1. Size of Scientific, Technical & Medical Information Segment Compared To Other Information Industry Segments

	Est. 2006 Revenue (\$M)	Est. 2007 Revenue (\$M)	Est. 2007 Growth (%)
B2B Trade Publishing	19,972	21,484	7.6%
Company Information	3,764	4,068	8.1%
Credit & Financial Information	36,495	40,492	11.0%
Education & Training	40,067	42,482	6.0%
HR Information	3,244	3,753	15.7%
Legal, Tax & Regulatory	12,633	13,905	10.1%
Market Research, Reports & Services	28,006	30,699	9.6%
IT & Telecom Research, Reports & Services	2,550	2,782	9.1%
News Providers & Publishers	134,671	133,300	-1.0%
Scientific, Technical & Medical Information	16,120	18,250	13.2%
Search, Aggregation & Syndication	32,980	41,550	26.0%
Yellow Pages & Directories	32,500	34,353	5.7%
Total Information Industry	363,002	387,118	6.6%

The questions have changed...

The transition from document-centric solutions

- What questions do you want to ask of data?
 - What is my competitor's IP Strategy?
 - Where am I duplicating efforts in my R&D organization?
 - What is the real-time customer perception of my new product?
 - What is the real-time customer perception of my market peers?
 - What are the emerging trends in my market?
 - How does my R&D compare with my competitors?
 - When and how will global events affect my stock price?
 - Where are the gaps in my consulting project portfolio?
 - Who are my top experts to compete for this RFP?
- None of these answers comes from a single document

MeSH - Disorders

[Kuru](#)
[Creutzfeldt-Jakob Syndrome](#)
[Slow Virus Diseases](#)
[Prion Diseases](#)
[Infection](#)
[Amyloidosis](#)
[Virus Diseases](#)
[Gerstmann-Straussler-Scheinker Disease](#)
[Animal Diseases](#)
[Strains](#)
[Cerebral Amyloid Angiopathy](#)
[Senile Plaques](#)
[Down Syndrome](#)
[Disease Models, Animal](#)
[Nerve Degeneration](#)
[Gerstmann Syndrome](#)
[Encephalopathy, Bovine Spongiform](#)
[Central Nervous System Diseases](#)
[Neuroaxonal Dystrophies](#)
[Neurodegenerative Diseases](#)
[Multiple Sclerosis](#)
[Central Nervous System Infections](#)
[Spinal Cord Diseases](#)
[Sheep Diseases](#)

Application

[direct methanol fuel cell](#)
[hybrid](#)
[polymer electrolyte membrane fuel cell](#)
[solid-oxide fuel cell](#)
[portable application](#)
[fuel cell system](#)
[electric vehicle](#)
[automobile](#)
[hydrogen use](#)
[automotive application](#)

Catalyst

[platinum](#)
[nanoparticle](#)
[selectivity](#)
[activation](#)
[precursor](#)
[titanium dioxide](#)
[sputter](#)
[surface modification](#)
[supported catalyst](#)
[nickel oxide](#)
[magnesium](#)
[impregnation](#)
[metal oxide](#)
[palladium](#)
[adsorption](#)

Mathematical Modeling/Analysis

[simulation](#)
[dynamics](#)

powered by
Collexis

Current thesaurus:

thesaurus Defence and Terrorism

iraqi	100	
bomb making	73	
iraq	70	
Ramadi	68	
weapons cache	65	
explosives	52	
car bomb	52	
rockets	51	
terrorists	46	
vehicle	38	
submachine gun	37	
rifle	36	
mosque	36	
mortar	35	
S. C.	35	
Baghdad	34	
grenades	34	

sulfonation level

polystyrene sulfonic acid

Expert Profiles

Prof. Dr. med. Wolfgang R. Lanksch
International Neuroscience Institute Hannover: Neurochirurgie

Navigation Menu (Left):

- Forschungsprojekte
- Krankheiten und
- chronisches
- News
- Introduction
- Academics
- Administration
- Admission
- Scholarships
- Exchange Programs
- Academic Projects
- Bulletin
- Banner

Header (Top):

JOHNS HOPKINS UNIVERSITY
NATIONAL INSTITUTES OF HEALTH
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Navigation (Top):

Home Categorize Document Show DC Modify DC Approve DC DC Analysis Budget Reports On Demand Logout

Main Content (Mayo Clinic Dashboard):

Modify Disease
Disease Category: <http://red.com>

Comparing D to Grant '5 R IDU'S'

overlapping Hepatitis has 7 c

concept name

- Hepatitis B
- Hepatitis B Vacc
- Hepatitis
- Hepatitis B Incid
- Hepatitis B Vacc
- Hepatitis Viruse
- intravenous drug

concepts in FP
*FP 1 = Hepatitis (D)

non overlap Hepatitis has 201 EXCHANGE BASED

concept name

- 16 kDa protein, hep
- 21 kDa protein, hepatitis C virus

Angulo-Hernandez P 95 Publications

Publication Timeline

Year	Count
< 96	1
96	9
97	8
98	12
99	14
00	10
01	12
02	18
03	7
04	4
05	1
06	1
07	1

Profiles Publications Experts Journals Departments

MeSH

- Disorders
- Cholangitis, Sclerosing
- Biliary Cirrhosis, Primary
- Fatty Liver
- Liver Diseases
- Fatty Liver, Alcoholic

Chemicals

- Ursodeoxycholic Acid
- Cholagogues and Choloretics
- Aspartate Aminotransferases
- Alkaline Phosphatase
- Biological Markers
- Alanine Transaminase

Collexis

Grants – examples of new questions

The screenshot shows the website for the National Center for Research Resources (NCRR), part of the National Institutes of Health. The page is titled "CTSA Funding Guidelines" and provides information about Clinical and Translational Science Awards (CTSAs). The navigation bar includes links for Home, Health, and Grants. The main content area features a "Quick Links" sidebar, a breadcrumb trail, and a list of funding opportunities. A "TAKE NOTE" section highlights key information about the 2008 funding strategies and peer review approach. The page also includes a search bar and a "GO" button.

U.S. Department of Health & Human Services
www.hhs.gov
Employee Info | Staff Directory | En Español

National Institutes of Health
The Nation's Medical Research Agency

HOME HEALTH GRANTS

National Center for Research Resources
National Institutes of Health
Department of Health and Human Services

SEARCH NCRR: GO

CHANGE TEXT SIZE: S M L

HOME ABOUT US PUBLICATIONS RESEARCH FUNDING SCIENTIFIC RESOURCES NEWS & EVENTS CONTACT US

Quick Links

- A-Z Subject Index
- Advisory Council
- Funding Opportunities
- Job Opportunities
- Meeting Reports
- NCRR Programs
- Program Contacts
- Site Map
- Strategic Plan
- Upcoming Events
- Visitor Information
- What's New

NCRR Home > Clinical Research Resources > Clinical Scientific Resources > Science Awards > CTSA Funding Guidelines

CTSA Funding Guidelines

SEE ALSO: Clinical and Translational Science Awards • CTSA Funding Guidelines: Questions and Answers

On September 18, NCRR issued a [news release](#) announcing the 12 new **Clinical and Translational Science Awards (CTSAs)** made in FY 2007. The total cost of these 12 new awards, which are almost 5 years in duration, will be \$574 million.

In 2006 the NCRR funded the first cohort of 12 CTSA institutions. To meet the goal of funding a total of 60 CTSAs in 2012 with a budget of \$500 million per year, and based on experience gained over the past two cycles, NCRR has developed new funding guidelines that have already been shared with the new awardees. In general, the guidelines are as follows:

- New funds included in the CTSA award will provide an increase of 35 – 45 percent over pre-existing NIH clinical research resources (i.e., NCRR K12, K30, M01, and Roadmap T32 and K12), up to a ceiling of \$6 million total in new funding.
- Certain successful applications may have low levels of the pre-existing NIH clinical research resources. To ensure that these CTSAs are successful, NCRR will support a minimum total award of \$4 million, or \$1.5 million in new funding, whichever is the higher amount.

The NCRR has notified the 2006 CTSA awardee institutions that they must follow these guidelines when competing for renewal of their CTSA in FY 2011. The 2006 CTSA institutions that are currently receiving increases over their pre-CTSA program levels in excess of the new

CR QUICK LINKS

- Staff Contacts
- Program Areas
- Resource Directory
- Funding Opportunities
- Program Guidelines
- News & Events

TAKE NOTE

- CTSA Pre-submission Videocast - March 7, 2008, Available for Viewing
- FY 2008 Funding Strategies
- CTSA Applications - Peer Review Approach
- CTSA Funding Guidelines
- NCRR Review of Carryover Requests
 - CTSA Carryover Requests

7, 8, 9

What is possible?

- Search has become a commoditized
- Knowledge management has finally become a reality...
- Ideas can be actively mined and linked from documents, websites, blogs, etc.
- Documents can be compiled for aggregation, clustering, profiling, categorization, visualization
- Trends can be mapped from unstructured and semi-structured document sets
 - What effect does political news (unstructured) and weather data (structured) have on corn futures?
 - What are the three chemicals which have not yet been linked in the literature to a Alzheimer's, but which are likely targets for disease interaction/prevention?
- Information, inference, discovery can become highly personalized

Examples...

The screenshot displays a Windows desktop with several Internet Explorer browser windows open. The windows are:

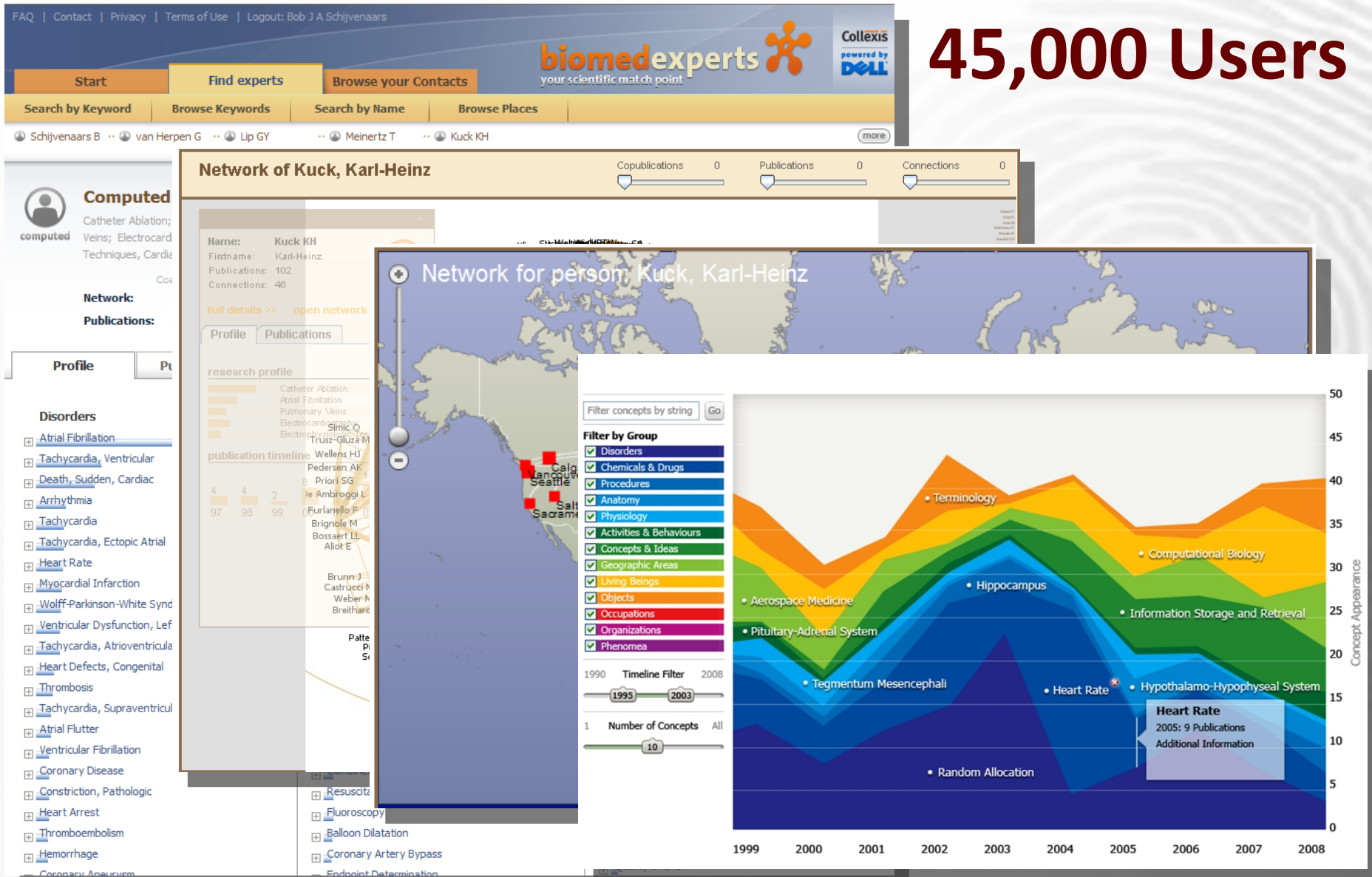
- Clusty Search**: <http://clusty.com/search?input-form=clusty-simple&v%3Asources=webplus&query=malaria>
- Grokker - Enterprise Search Management**: <http://live.grokker.com/grokker.html?query=malaria&Yahoo=true&Wikipedia=true&numResults=250>
- TEMIS text intelligence**: <http://www.temis.com/index.php?id=124&sel=1>
- evolvus.com**: <http://www.evolvus.com/di.htm>
- SciFinder - How to Explore by Exact Structure**: <http://www.cas.org/support/scifi/howto/structure.html>

The SciFinder window is the largest and shows a chemical structure drawing interface. It includes a toolbar with various drawing tools (Atom, Short, X, R, [] n, etc.) and a central drawing area. The drawing area contains a chemical structure of a phosphorus-containing molecule. The structure is a six-membered ring containing one oxygen atom and one phosphorus atom. The phosphorus atom is double-bonded to an oxygen atom and single-bonded to a nitrogen atom. The nitrogen atom is bonded to a carbon atom, which is bonded to a chlorine atom. The carbon atom is also bonded to another nitrogen atom, which is bonded to another carbon atom, which is bonded to another chlorine atom. The structure is shown in a perspective view.

The SciFinder window also displays a list of elements: C, H, O, S, N, P, Cl, Br, F, Si, I. Below the list are buttons for Preview, Get Substances, Get Reactions, and Cancel. The molecular formula is shown as C7 H15 Cl2 N2 O2 P and the weight is 261.09.

Biomedexperts.com – our first Network

45,000 Users



What is the portfolio?

- **Stop thinking about it as a set of documents and sites, start thinking about portfolios of information...**
- **Which projects are most profitable (or likely to succeed)**
 - How can I predict that in advance of the project
- **Which microsites are promoting my new drug?**
- **Are there adverse event trends on the web for my drug candidate?**
- **What other questions can I ask from my information portfolio?**
 - Which ones are worthy of the investment to answer them?
 - ...in real time
 - ...at any future point that I want to ask...

The new limitations – USERS *(and interfaces...)*

- The capabilities of knowledge management have far eclipsed the understanding of even data-intense users
- The capabilities of text mining have not even been born into the minds of most corporate executives
- How do we take this limitless capability and "iPhone" it into 'user-guide free' applications
- How do we get leadership buy-in without glazing them over with the "implications of OWL in a Web 2.0 world."

Key Messages...

- **For the tech buyers - If you can dream it, 'we' can do it**
 - Let 'us' do it / build it
 - Think about competitive intelligence investments over time
 - The 2 rate limiting factors – your imagination and your budget
- **For the tech sellers – GUI, GUI, GUI**
 - Think about design and UI as a competitive advantage
 - But not one that is likely to last...
 - Become a first mover
 - INVEST in training/education of your users
 - Make mistakes and apologize quickly...
- **For the big players – buy, partner, or perish**

Questions

The logo for Collexis features the word "Collexis" in a bold, black, sans-serif font. Above the "x" is a horizontal orange bar with a white circle in the center, resembling a stylized "C" or a fingerprint. The background of the slide has a faint, large fingerprint pattern.

Collexis

www.collexis.com

Stephen A. Leicht
leicht@collexis.com

803.727.1113