# SchemaLogic Inc.



*SchemaLogic provides a system to prevent taxonomy creep. The SchemaLogic system ensures a common tagging vocabulary across an organization via a separate content management system. In 2011, the company was acquired by Smartlogic.*

*Author's note: I originally prepared this report in 2004 for a firm interested in taxonomy software. I revised the client report in 2008 for an Outsell/Gilbane monograph. This is an unpublished, preliminary draft. The last substantive editing cycle was in late 2007. SchemaLogic struggled and was purchased by SmartLogic, a competitor in the enterprise taxonomy sector.*

*SchemaLogic was one of the first companies to build a system that provided a licensee with a way to organize controlled terms in a manner that enforced ANSI standards for a controlled vocabulary. The approach was to graft a traditional content management system with a classical taxonomy method. When the company was founded, most organizations did not have formal rules for indexing. The taxonomy craze rose and then fell. SchemaLogic's software proved to be a tough sell to companies that did not understand the value of consistent indexing for enterprise content.*

*SchemaLogic built a following among a select group of organizations. The one characteristic these customers shared was an understanding of the importance of assigning index terms for a single, centralized, fresh controlled term list. The organizations also had money to invest in a sophisticated system that requires attention from trained specialists. SchemaLogic's system was one of the first to enhance the value of information managed in a SharePoint system.*

*The challenge was closing deals and generating sufficient cash to pay back the investors. An innovative product is not necessarily successful. product. This information is a rough draft and is frozen.*

*Stephen E Arnold, November 27, 2013*

# Introduction

SchemaLogic provides a metadata management system. Experts in indexing know that uncontrolled vocabularies make finding information difficult and frustrating. Taxonomy creep means that indexing is uncontrolled, subject to the whims of those adding tags. SmartLogic built its business on selling software that ensures that consistent metadata are used by separate enterprise software systems.

Table 1: SchemaLogic Enterprise Suite: A Bird's Eye View

| Product Thumbnail | |
|---|---|
| 1 Search Brand | SchemaLogic Schema Server |
| 2 OS Supported | Microsoft Windows running SharePoint |
| 3 Est License Fee | $150,000. Custom price quotation required. License fees do not include consulting and engineering services. |
| 4 Functions | A content management system for controlled terms used to index enterprise content. The system includes Schema Server, an Integration Service, Real-Time Semantic Services (RTSS), Collaboration Services and the Workshop for Web access |
| 5 Claimed Features | The system works in a manner analogous to an "air traffic control" service. |
| 6 Downsides | Setting up the system requires scripting to knit together the systems that will be using the SchemaLogic Server. A subject matter expert (index specialist) is required to manage the term lists. The SchemaLogic system is designed for customers who understand and value a coordinated, centralized, and managed approach to metadata. |
| 7 Similar To | The original Cuadra STAR system (now a unit of LucidIdea and Mondeca) and library-centric systems like Access Innovations' controlled term system. |
| Product Close Up  | SchemaLogic provides software that automates the link between a controlled vocabulary and enterprise software like Microsoft Office and SharePoint. The system makes sense for organizations that have a commitment to rigorous, consistent indexing. SchemaLogic's approach eliminates the manual hassles of using the same metadata across disparate systems.The workflow and integrator components of the SchemaLogic solution provide process and technology that enables IT workers to synchronize of metadata across different enterprise systems. For organizations without an understanding of consistent indexing, SchemaLogic is software is a tough product to sell. |

SchemaLogic was founded in 2003 by Trevor Traina and Breanna Anderson. Before getting involved with metadata management, Mr. Traina worked at Compare.Net and sold that company to Microsoft in 1999. Mr. Traina enlisted Breanna Anderson, a Microsoft expert and launched SchemaLogic. Prior to SchemaLogic, Ms. Anderson was a software architect and program manager at Microsoft from 1995 to 2001.

SchemaLogic is notable because the company is taking a classical approach to taxonomy management via a content management or CMS architecture. Its approach to metadata provided a solution to the often chaotic indexing applied by users of SharePoint and by automated indexing systems integrated into third-party search systems. From its inception, the firm focused

on the licensees of Microsoft SharePoint, Microsoft's still-evolving content management and collaboration platform, and Microsoft enterprise search technologies. SchemaLogic provided functions that were not included in the SharePoint system.

In 2007, Breanna Anderson left SchemaLogic, where she had served as the company's chief technical officer for four years. She was the architect of SchemaLogic's complete suite of enterprise metadata management solutions and authored the firm's key patent "Schema Server Object Model," US 20040181544 in 2003.

The company rolled out its first product in November 2003, and in the last four years has refined the company's metadata content management technology. The SchemaLogic product remained mostly unchanged up to the time of the sale of assets to Smartlogic, a direct competitor.

## Funding: Big Numbers

Trevor Traina, the company's co-founder and co-chairman, provided initial capital for the venture. The SchemaLogic CEO was Jeff Dirks, who assisted Andrei Ovchinnikov in June 2003, prior to the "official" launch of the company.[1] In 2003, SchemaLogic's initial funding is estimated to be about $5.0 million from Seattle-based Phoenix Partners and several other investors. The company completed a $4.6 million series B round in December 2004, led by Seattle-based Madrona Venture Group. In March 2007, Mr. Dirks helped the company obtain an additional $14.7 million in financing led by Goldman Sachs with participation from Chevron Technology Ventures, Madrona Venture Partners, and Phoenix Partners. The March 2007 cash infusion pushed SchemaLogic's total funding into the $26.0 million range. SchemaLogic remained a privately held firm and did not report its revenues. This series C round of funding kept SchemaLogic in the game as metadata fever raged through the search sector from 2007 to 2009. It is not clear if SchemaLogic was able to hit its revenue targets.

The privately held firm does not report its revenues nor does it provide information about the firm's employees. (At the time of its sale to Smartlogic in September 2011, SchemaLogic had lost its market momentum.) In 2007, SeattlePI reported that the company had revenues of about $20 million.[2]

## SchemaLogic Management

- Trevor Traina, co-founder

---

1. Mr. Ovchinnikov is an entrepreneur and a martial arts expert.
2. John Cook, "SchemaLogic Inks Deal with the Associated Press," July 16, 2007 at http://blog.seattlepi.com/venture/2007/07/16/schemalogic-inks-deal-with-the-associated-press/

Enterprise search is becoming the de facto way of finding information in corporations today. By providing consistent metadata structure and meaning via aggregated taxonomies and controlled vocabularies, we enhance SharePoint and other systems' functionality. Text mining analyses gain an immediate boost in accuracy due to the use of standard metadata across all systems.—A SchemaLogic Executive

- Andrei Ovchinnikov, co-founder
- Jeff Dirks, chief executive officer and former Capital Stream executive
- Breanna Anderson, technology officer
- Lowell Anderson, vice president of marketing, who coined the phrase "high definition findability"
- Tom Pietryga, vice president of sales. He worked at Documentum, a system that SmartLogic supports
- Mark Glover, engineer
- Michael Doane, taxonomy specialist

## Customers

In 2007, SchemaLogic said that it had thirty-five licensees.[3] These firms understand the value of deploying a management indexing system that is complete, general, granular, consensus managed, culturally adaptable, consumer oriented, actionable, and evolutionary.[4] How many firms who value these attributes want to buy a metadata management system?

Licensees of the SchemaLogic system include:

- Associated Press. (Note: The AP licensed the InfoDesk system to perform some content processing for its owners. The AP uses the Teragram system for entity extraction and automated indexing.)
- Boeing
- Chevron. The company appears in some company literature as an "investor."
- Church of the Latter Day Saints
- Corbis. Indexing applied to images
- IBM
- Kelly Search (Reed Business Information)
- Pfizer
- Procter & Gamble
- Raytheon.

3. See http://blog.seattlepi.com/venture/2007/07/16/schemalogic-inks-deal-with-the-associated-press/

4. Source of the list is a lecture given by Breanna Anderson, "Growing the Tree of Agreement: Building, Managing, and Disseminating Metadata Standards through Online Collaboration," February 20, 2004 at Online Northwest.

We are seeing that the trend within the top analyst firms is to talk more specifically about business semantics management, search, enterprise content management, and collaboration which are really subcomponents to the broader information management market. Of course, from the SchemaLogic standpoint, we believe that business semantics management is a cornerstone to solving this broader set of issues around information management —Jeff Dirks, SchemaLogic at http://www.twst.com/interview/24478

## Partners

Since its inception, SchemaLogic operated as a Microsoft partner and then in 2008, SchemaLogic inked a deal to become an IBM partner. I interpreted this announcement as an indication that SchemaLogic was seeking additional revenue opportunities. The company also signed on to be a partner of Concept Interactive and Vamosa, a "content governance" company[5] as well as Advent One in order to extend the firm's reach in the Pacific Rim.[6] The company tied up with Meta Integration Technology. SchemaLogic offers connectors for databases and enterprise applications such as SAS, Oracle and IBM Rational.

Other partners are:

- Bimodality Isogen - Innodata Isogen is a publisher service bureau that provides rekeying, indexing, and taxonomy development. SchemaLogic teamed with Innodata in 2004 to offer metadata management technology. The companies have been working together to assist their customers with the management and aggregation of XML content.
- Intellisophic - Intellisophic develops taxonomies for enterprise clients and offers prebuilt taxonomies to its clients. The deal between SchemaLogic and Intellisophic allows SchemaLogic to deliver to its customers pre-built taxonomies. The deal acknowledges that customers want integrated solutions for metadata. SchemaLogic can import Intellisophic taxonomies. The customer uses the SchemaLogic tools to customize and maintain the word lists.
- Siderean - this vendor of semantic search systems has inked a deal with SchemaLogic "to help organizations improve the 'findability' of relevant information across their enterprises." Siderean's approach to search requires extensive metadata to support its faceted display of categories and "see also" links.
- WAND - this is another taxonomy development shop for vertical markets. The two companies linked up to deliver "a comprehensive taxonomy solution to Global 2000 companies." The partnership offers customers taxonomic content, coupled with a platform to view, manage, and disseminate the taxonomies to their search and content management applications. The idea is to deploy a "ready to run" taxonomy solution instead of investing in a time-consuming taxonomy development project.

5. See http://www.it-director.com/technology/news_release.php?rel=12220. "Content governance" may be marketing-speak for editorial policies and enforcement.
6. See http://www.mobilitytechzone.com/news/2006/04/20/1590766.htm.

## What SchemaLogic Does

The problem SchemaLogic "solves" is the different versions of metadata that separate systems create and use. The core is a "schema" for a controlled term list used to index content in an organization.

SchemaLogic, according to the company's Web site asserts that it:

> provides the only enterprise scalable platform and collaboration and management of a metadata plan, installation of the plan in SharePoint, and synchronization of the evolving metadata throughout the global SharePoint environment.

Users typically enter queries in free text, but different enterprise software applies different indexing terms to content. Users often become accustomed to using certain words and phrases to locate information. If the indexing is not consistent, users find that locating the required information becomes more and more difficult due to indexing "drift."

SchemaLogic captures and communicates controlled terms used across different applications. The software simplifies information about documents and the terms used to index those documents. The SchemaLogic system imports existing taxonomies. controlled term lists, and reference data. The system can process content in relational databases, some applications like Microsoft Office, and XML schema (if in use at the licensee's organization). These metadata are stored in the SchemaLogic service that helps ensure that consistent indexing is applied across the different systems.

The Schema Server then manages the associations and dependencies among the different schema. A subject matter expert can use the editing interface to map See Also, Use For, and other relationships among index terms and entities. Other enterprise systems "subscribe" to the Schema Server in order to synchronize new or update controlled terms and taxonomy entities. The interaction with enterprise applications and the Schema Server makes use of Extensible Markup Language, Web services, and specifications like Simple Object Access Protocol.

Table 2: What Metatagging SchemaLogic Can Support

| Method | Standard |
|---|---|
| 1 Encoding | Standard Generalized Markup Language, Electronic Data Interchange, Extensible Markup Language, ASCII flat file |
| 2 Metadata schema | Document Type Definition, Extensible Markup Language Schema, External Data Representation, REgular LAnguage for XML Next Generation, Resource Description Format, ISO/IEC 11179 (formally known as the ISO/IEC 11179 Metadata Registry (MDR) standard), Unified Modeling Language, Electronic Data Interchange, Meta Object Facility, OWL Web Ontology Language, and XML Metadata Interchange. |
| 3 Taxonomy Standards | ANSI/NISO, z39.19, OWL Web Ontology Language, Resource Description Format, Topic Maps, ISO 11179 |

SchemaLogic pricing: “Enterprise deployments range from $100K through $1M based on the size of the deployment.”—Jeff Dirks, Vice President, Marketing at http://www.twst.com/interview/24478

| Method | Standard |
|---|---|
| 4 Service Definition/Execution | Web Service Definition Language, Simple Object Access Protocol, Resource Description Format, Common Object Request Broker Architecture, Java Remote Method Invocation, Distributed Component Object Model |
| 5 Directory/Discovery | Universal Description, Discovery and Integration, Simple Object Access Protocol, Lightweight Directory Access Protocol, Microsoft Active Directory, Novell Directory Service |
| 6 Metadata/Schema Standards | Dublin core, Defense Information Systems Agency XML Emporium, Metadata Encoding and Transmission Standard, MPEG-7 Image Metadata Standard, Really Simple Syndication, Prism Metadata Initiative, Adobe’s Extensible metadata Platform, Information and Content Exchange, Electronic Business Extensible Markup Language, and OASIS Metadata. |

For example, in an organization with SharePoint, Microsoft Office, and a content management system like Documentum, SchemaLogic provides a unified information management system. Documentum manages metadata as part of its “managed change process”. SharePoint uses a more relaxed and dynamic approach to metadata. SchemaLogic provides a mechanism to perform metadata mapping between the SharePoint and the Documentum metadata. SchemaLogic offers software that allows an individual worker to generate content that contains indexing consistent with the controlled term list residing on the SchemaLogic server.

As the company ran into financial headwinds, the description of the company shifted from controlled terms to semantic jargon; for example:

> SchemaLogic provides Business Semantics Management (BSM) software for companies to manage and connect the different vocabularies used to describe information, aligning content management and search systems so people can find and share that information across their enterprise.[7]

## Products

SchemaLogic, Inc. provides metadata management software solutions. It offers MetaPoint, a solution for automating the link between Microsoft Office and SharePoint; and Schemalogic Enterprise Suite that enables business subject matter experts and IT professionals to define and manage a semantic standard. The company’s MetaPoint products include MetaPoint Desktop, a Microsoft Office add-in that recognizes and applies keywords and tags in a document, and suggests or assigns content types and document libraries; MetaPoint Server, which provides a centralized metadata management repository that supports the creation and management of a master

7. See http://www.techrepublic.com/blog/it-news-digest/ceo-jeff-dirks-says-schemalogic-helps-organizations-manage-their-information/

SchemaLogic solves the problem Bella Hass Weinberg, professor of library science, St. John's University, identified: "There is a singular lack of vocabulary control in the field of controlled vocabularies."—Source, Breanna Anderson lecture, February 2004

metadata model; and pre-built industry-specific vocabularies. Its Schemalogic Enterprise Suite comprises Schema Server, an object modeling, collaboration, and governance engine that supports simple list structures, complex multi-faceted taxonomies or ontologies, and enterprise-wide metadata structures that describe corporate information assets; SchemaLogic Connector for Microsoft SharePoint 2007 to manage SharePoint metadata; SchemaLogic Connector for EMC Documentum to manage Documentum metadata; and SchemaLogic Connector for Fast Search ESP to improve information access across the organization. The company also provides advisory services, including enterprise taxonomy assessments, solution design and data modeling, best practice workshops; and implementation, training, and technical support services. Its solutions are used for automated desktop tagging, intelligent content framework, and enterprise content governance applications by businesses in digital media, life sciences, and oil and gas industries.

The core product is the SchemaLogic Enterprise Suite. It offers a framework that provides the functionality required to map and manage semantic standards underpinning controlled term lists, taxonomies, and knowledge base content.

The system implements what I consider a rigid, top-down approach to controlling index terms and their use in an organization. Once a list is developed, the SchemaLogic system acts as a central repository of index terms and associated forms such as "use for" and "see also" tags. The taxonomy makes possible synonym expansion for indexing and query processing.

Schema Server is the active enterprise governance repository. Think of it as a content management system for metadata. The Schema Server feeds the normalized metadata to other enterprise software systems requiring metadata. In a sense, the Schema Server is a subject matter expert / editor and work flow engine for metadata.

The SchemaLogic Workshop is the desktop graphical application. Authorized users can manipulate the knowledge base and interact with the various mapping functions used in the system. SchemaLogic also offers a Workshop Web product. This browser-based tool allows users to view, manage, and collaborate to develop what SchemaLogic calls "enterprise-wide semantic models" or the rules the system applies to metadata. The idea is that users of the system are in the best position to enter a new term or modify a concept mapping. The SchemaLogic system handles the updating across the various enterprise systems tapping into the metadata repository, thus metadata are "in sync" without further manual intervention.

In 2007, the company added a solution for media, iSchemaLogic's controls could be applied to indexing of videos and images. However, nStein, later acquired by OpenText, the Canadian enterprise solution provider, demonstrated that a market existed for organizations struggling with video and other non-text content. The product was basic SchemaLogic applied to indexing images.

## How the System Works

The operation of the system is similar in some ways to content management systems. With SchemaLogic, organizations create and manage taxonomies centrally and capture multiple perspectives by mapping back to the central definitions. For example, different departments might use IBM, I.B.M. and IBM Corp - these variants will resolve to IBM Corporation, and pull all relevant information.

The “models” used by SchemaLogic are a combination of rules and term lists. The system can handle tagging sales regions as well as complex multi-faceted taxonomies with thousands of terms.

Schema Server describes the structural models used to store and exchange information as a hierarchy of information classes. This logical modeling capability allows a licensee to capture a consistent, easy-to-understand model of the information systems in the enterprise. The model specifies how these systems interrelate with each other. The system accommodates:

- Relational models; that is, traditional database schema
- Object-oriented models; that is, a class such as *Products* is used as a framework to instantiate an instance of a class such as a specific product.
- XML; that is, structured documents
- Service-oriented architectures; that is, metadata applied to Web services description language.

How these different models’ metadata are applied and what metadata to use are functions handled within the SchemaLogic system.

One way to think of SchemaLogic is to visualize the system as what Ms. Anderson calls “a digital metadata librarian.” The difference is that the SchemaLogic server automates most of the time-consuming work needed to keep metadata synchronized across different enterprise software systems. To get around the bottleneck of manual updates, SchemaLogic allows users to make adjustments to the term lists, thus reducing the cost of maintaining the knowledge base.

Keep in mind that the SchemaLogic system must be configured, its basic rules tweaked for your specific organizational requirements, and the workflow and other rules must be tweaked. Once set up, the SchemaLogic system can operate silently and with modest manual intervention and tuning.

## Examples of the System in Use

SchemaLogic’s customers include the Associated Press, Boeing, and Chevron, among others. SchemaLogic guards the identify of its clients closely. One client, Chevron, is an investor in the company. The Associated Press is

"Given the myriad of free content across the Web, it is no secret that the media and publishing market has gone through a tremendous transformation over the past decade. These leading organizations know they need to optimize their information syndication services and aggressively take a hard look at subscription revenues. Through the use of enterprise-wide Business Semantics Management solutions, they are able to aggressively compete with the multitude of open source content providers which ultimately results in accelerated revenue."—Jeff Dirks, Schemalogic CEO. Source: Information Today at http://goo.gl/GcrpHQ

an example of a publishing company with a need to standardize indexing of the content disseminated to its newspaper owners and licensees.

### *Associated Press*

The Associated Press uses SchemaLogic to index its content in five languages and repurpose its information. In addition to classifying news stories consistently, the AP relies on SchemaLogic to reduce indexing and knowledge base maintenance costs. The AP is an investor in InfoDesk, a content processing and online distribution service. The AP has integrated SchemaLogic with various third-party services used to serve the syndication owners and content licensees. InfoDesk includes a content classification and indexing function in its service.[8]

### *Boeing*

Boeing uses SchemaLogic server to create a central set of definitions that describe their products, goods, and services that can be centrally managed by the company's library services organization. The use of trained librarians provides the indexing expertise essential to the effective use of the SchemaLogic system. Boeing units "subscribe" to the SchemaLogic index management system's knowledge bases. The payoff is that a single set of indexing terms and metadata are applied across different business units. The system makes it possible to assign consistent indexing to engineering drawings and images. My understanding is that Boeing uses SchemaLogic to integrate three types of metadata:

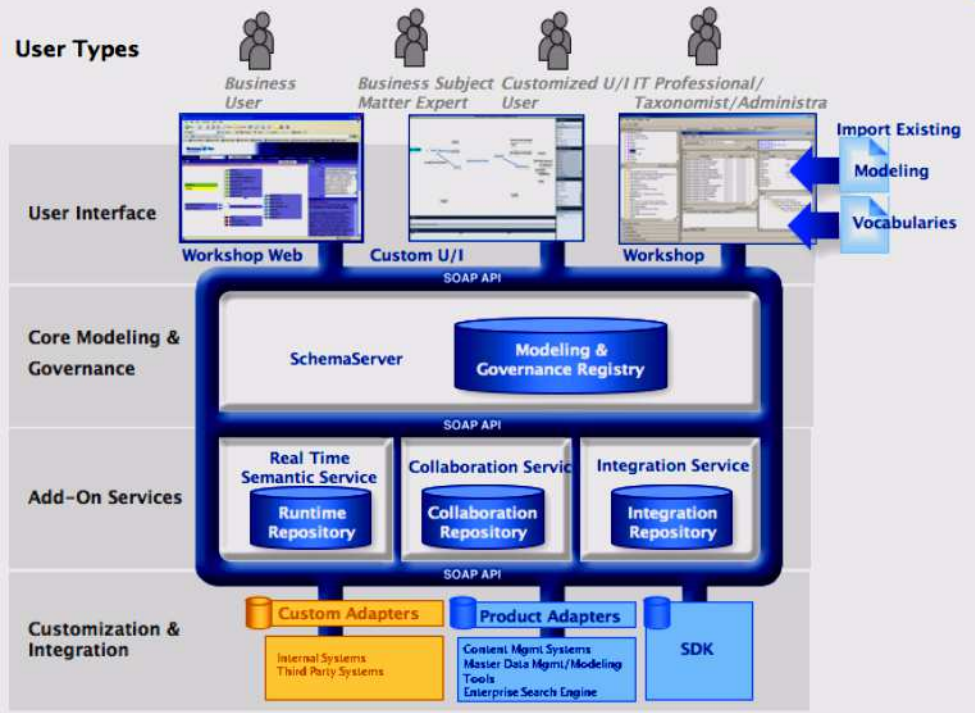
- Technical metadata; for example, standards, material science terminology, and Boeing jargon
- Business-related metadata; for example, vendor and customer company names, vendor identification numbers, and purchase information
- Unstructured metadata; for example, file names, date and time stamps, and file types.

### *Chevron*

Chevron, a SchemaLogic investor, uses SchemaLogic in its Global Information Link initiative. Chevron has deployed Microsoft Office SharePoint Server 2007 as its core search engine. SchemaLogic makes it possible that

8. See www.infodesk.com

Chevron's digital information assets are described in a consistent way globally across Chevron's many operations.




This SchemaLogic diagram shows the system's four key features: [a] a customization component via the APIs and configurable adapters, [b] the core modeling and workflow server and [c] its add-on services, and [d] the customizable interfaces to the Web interface, a customized client interface, and the "workshop" for managing the knowledge bases.

The SchemaLogic system is used to synchronize Microsoft SharePoint installations within the Chevron organization of more than 56,000 employees.

### *IBM*

IBM, a SchemaLogic partner, uses SchemaLogic to manage its metadata for both structured and unstructured information. The controlled terms make it possible for IBM to apply consistent indexing across its products and services, business units, and systems. Applications of the SchemaLogic Server include:

- Indexing content for the OmniFind search system and third-party search systems
- IBM's internal XML repository for technical reports, standards documents, and documentation

- Unstructured data such as contracts, marketing information, and email.

After installing the SchemaLogic system, IBM expanded the role of the SchemaLogic system to include:

- Staff expertise and business unit competencies
- Technical information
- Proprietary IBM internal systems; for example, the On Demand Workplace.

IBM does not reveal return on investment associated with the SchemaLogic solution, but the company seems to be committed to SchemaLogic's managed metadata system.

### *Procter & Gamble*

The consumer product giant wants to apply corporate standard brand metadata to the company's business units and subsidiaries that operate in many countries. Consistent metadata helps Procter & Gamble build and protect brand identify. SchemaLogic makes it possible for a P&G professional to add a product name or a product category as needed. Other P&G professionals have access to these new categories quickly. Because Systems from EMC Documentum, Oracle, and other enterprise systems, a P&G professional is able to get access to current information quickly. The idea is for P&G to reduce the time required to locate information about new products or market developments.

Proctor & Gamble suggests that SchemaLogic has reduced costs. These savings have come from the elimination of the headcount required for manual editing of index terms. P&G, like a pharmaceutical company, operates within a Web of regulations. If a legal matter arises, P&G can locate information using a consistent set of classifications, controlled terms, and categories. Does the SchemaLogic system reduce risk? P&G believes this assumption to be accurate.

## Technology

SchemaLogic's server is a centralized repository for defining and maintaining content type metadata definitions. It also houses list values that can be distributed and monitored for compliance across the distribution content systems, data farms, and site collections in an organization.

Schema Server allows information architects to assign semantic relationships (such as "author of," "related to," "component of," "skills needed for"

and “cause of”) and other descriptions-from a system-independent perspective- making it easier to mine and analyze the most relevant information.

The repository stores both the schema, usage data, and rules. The server differentiates between what it calls structural schema and taxonomic schema.

Schema Server describes the structural models used to store and exchange information as a hierarchy of information classes. Schema Server is a database and content management system that manipulates:

- Content classes
- Elements
- Vocabularies
- Terms
- Vocabulary views.

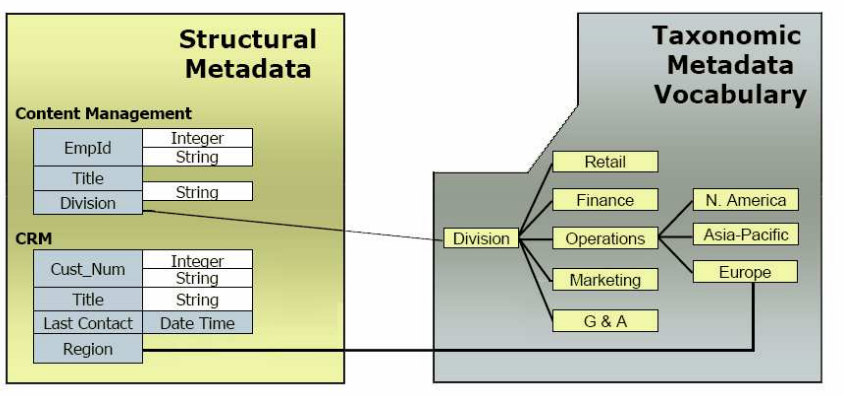
Structural schema are metadata that identify a file, such as name, date, size and so on. The taxonomic schema refers to metadata that describes what the content of the file is about.

SchemaLogic’s spider and indexing system can be controlled by the system administrator in order to provide the optimal indexing for a particular client installation. SchemaLogic, for example, provides an administrative setting with which the depth of the crawl and metadata processing can be controlled. Note that SchemaLogic’s initial processing can consume significant CPU cycles and days or weeks of CPU time. However, incremental updates occur more rapidly.

The server stores data for the workflow functions. In the SchemaLogic context, *workflow* means enabling users to model relationships between data, to map associations among objects, and to resolve inconsistencies in the schema and among objects. Schema Server allows the system administrator to set up approval mechanisms so that certain changes require the approval of the manager of a particular schema, taxonomy, or controlled term list.

SchemaLogic provides administrative tools to manipulate knowledge bases and “lists”. A change to the knowledge bases ultimately has an impact on what word or phrase is assigned to a category. Authorized users manipulate these metadata, not the documents or the index terms assigned to a document. A SchemaLogic user, for example, can add concepts to a schema’s entry. These schema entries are then applied to the documents or other information objects. Once a change has been made in the Schema Server, that change is available to other systems with access to the metadata. Schema Server can also update indexes for those third-party systems for which adaptors exist. A licensee can create

additional adaptors to handle applications not supported by the adaptors that are included with the Schema Server.




The structural metadata are mapped to taxonomy metadata. The approach makes extensive use of standard relational data tables.

## Integrator Server and Adaptors

SchemaLogic has adapters for Autonomy Inc. (IDOL and K2), BEA Systems WebLogic portal, EMC Documentum, FileNet, IBM WebSphere portal, IBM WebSphere Information Integrator Content Edition (the part that includes the acquired Venetica technology), and Oracle Stellent.

SchemaLogic's "integrator" technology facilitates taxonomy mapping and synchronization across enterprise systems-search engines, automatic categorization, content management, and databases. The core of the SchemaLogic Integrator is the integration server. This server manages taxonomy synchronization.

Adaptors run in the integration server and communicate directly with subscribing systems. The adapters "watch" the database in the repository for changes. Only documents known to the system can be updated.

The integrator and adaptor combination synchronizes taxonomy data, linguistic information, and controlled vocabularies between the application and Schema Server's central repository. Each of SchemaLogic's adaptors work in similar ways for Vignette, EMC Documentum, and the other supported systems.

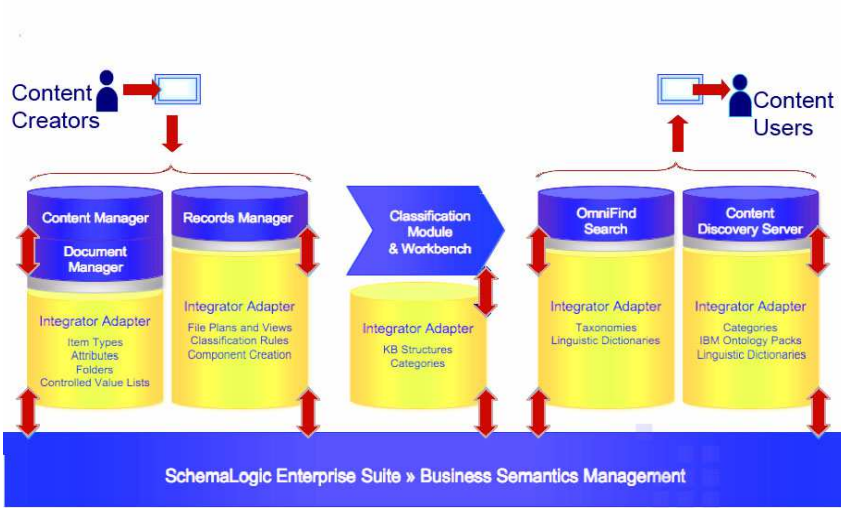
The IBM adaptor is representative of the SchemaLogic approach. The IBM adaptor consists of several components. These are:

- OmniFind Rule-based Taxonomies - the OmniFind Integrator Adapter publishes enterprise taxonomies and category rules to OmniFind. These

taxonomies are used by OmniFind to enable classification and to enhance search and content browsing.

- Synonym Dictionaries - custom Synonym Dictionaries can be generated from the enterprise taxonomy, allowing users to expand their search results to include content that contains synonyms of query terms. The synonym dictionaries support “See also” and suggestions about other topics germane to the user’s query.
- A Term List to Boost Certain Metadata - a *boost word* is a term that appears higher in a results list or is in some way flagged as featured content. A boost factor can be applied to a term. The numeric value increases or decreases the term’s importance.
- Stop Word Dictionaries - custom Stop Word Dictionaries can be generated from the enterprise taxonomy, identifying terms to be removed from search queries in order to increase search relevancy. (A stop word is one that the indexing system skips; for example, *and*, *or*, *but*, etc. )

The diagram shows the major components of the IBM integrator adaptor. SchemaLogic customers can install one or more of the adaptor modules, perform necessary customization, and manage the system from SchemaLogic’s graphical administrative interface.

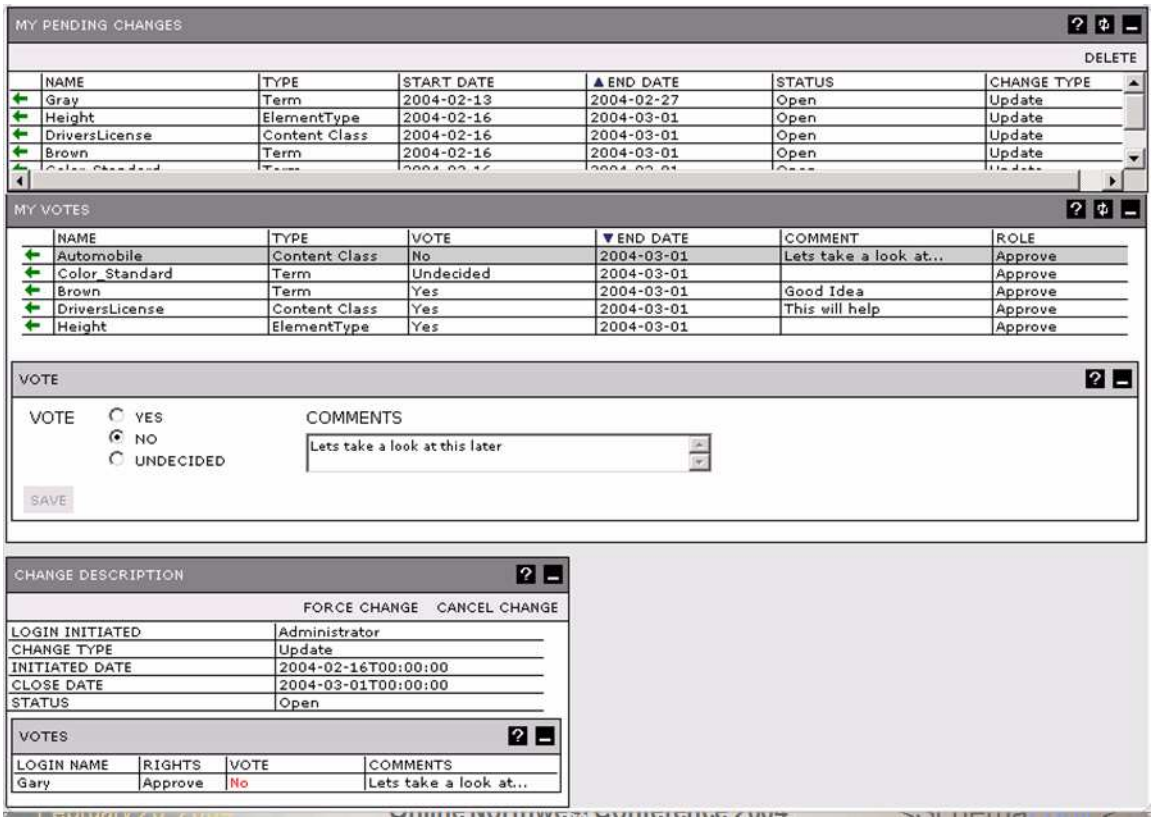



The SchemaLogic system provides an enterprise-wide knowledge base. Other enterprise applications tap into the SchemaLogic taxonomy repository. A user will be able to use standardized concepts regardless of the application exposing concepts and terms.

In addition to the adapters, SchemaLogic offers for IBM installations IBM Ontology Packs. These are canned controlled vocabulary lists. Word lists are

available for finance and pharmaceuticals. Additional vocabularies can be licensed from WAND, a SchemaLogic partner. One of SchemaLogic's semantic features is a "grammar list." The administrative interface makes it possible to manage linguistic dictionaries for synonyms and for lemmatization (stemming).

SchemaLogic also provides a component that can integrate with an IBM's classification service. A user of an IBM's discovery solution can apply consistent classifications and index terms across IBM enterprise solutions.




The graphical interface for controlled term and taxonomy management is designed for a person with expertise in metadata. The SchemaLogic system may require a system administrator with experience in taxonomies and indexing. If the system administrator lacks these skills, then the licensing organization may require on a full-time or contract basis, one or more professionals who interact with the metadata management system.

Licensees can use individual components or mix and match them to meet specific requirements. SchemaLogic provides the software needed to support IBM content management systems without any custom coding.

With the SchemaLogic server and adapters up and properly resourced, a licensee can establish and enforce a company-wide information model, stan-

dardized enterprise taxonomy, and controlled vocabularies. For a company with an IBM Web Sphere Information Integrator OmniFind Edition and Web Sphere Content Discovery Server, SchemaLogic Server enables:

- Linking terminology across disparate content repositories for index consistency
- Decreasing the error rate of automatic classification operations
- Supporting faceted search/See Also inclusions
- Implementing Endeca-like or Siderean-type navigational links.

# Selected SchemaLogic Features

The SchemaLogic Server includes a number of interesting features. Some of these respond to the expected interest in collaborative indexing; that is, professionals with access to document creation functions can assign their own index terms to the content object. The idea SchemaLogic implements is that index conflicts can be resolved by users. Users can also add terms and bound phrases to the master controlled term list. Other ideas are designed to allow licensees to integrate with third-party enterprise applications or create specialized functions to handle operations not native to the SchemaLogic Server.

SchemaLogic supports most Java-compliant environments. The current release adds certification for AIX, HP-UX, and Sun operating systems to the existing support for Microsoft's operating system. SchemaLogic supports a variety of Application servers, including IBM WebSphere and BEA WebLogic. Databases certified for use with SchemaLogic include IBM DB2, Oracle, and SQL Server. Structural metadata is expressed in XSD, shorthand for XML schema definition files.

## Collaboration to Resolve Metadata Clashes

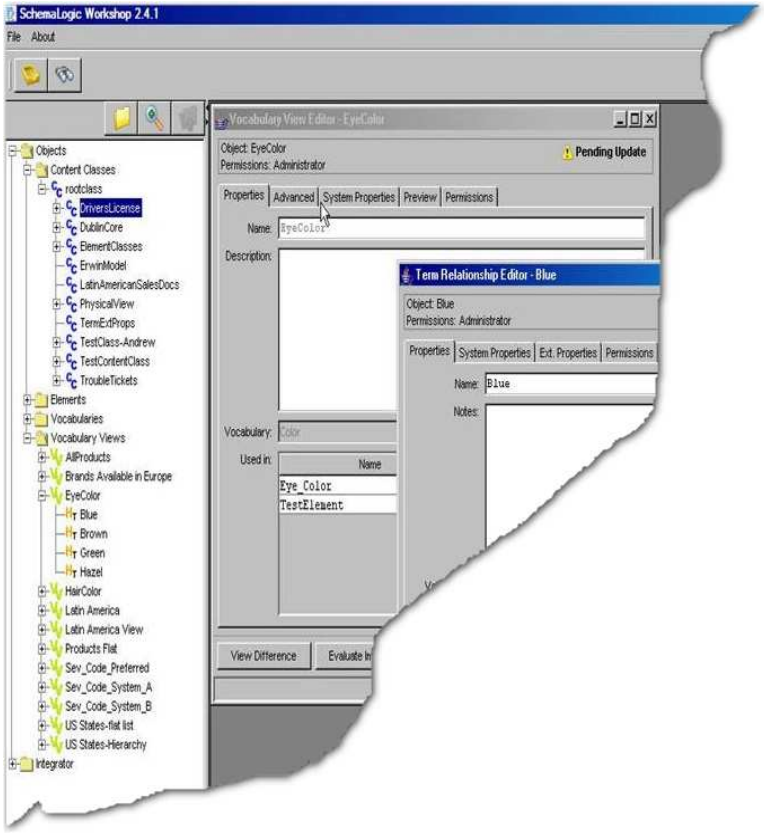
SchemaLogic employs a partner program to meet the needs of its customers. SchemaLogic offers consultants and resellers a partner program. Partners receive access to technology and training from SchemaLogic. SchemaLogic and its partners cooperate with marketing the SchemaLogic server. The idea is that partners contribute specialized expertise in vertical markets, content management, XML, data integration, or information architecture to design and implement solutions using SchemaLogic technology. SchemaLogic partners include Microsoft, IBM, EMC2/Documentum, Fast and Metalogix.

Jeff Dirks said in Dan Keldsen's Web log in 2006:

> We really see collaboration and participation at the heart of what we are calling business semantics management and frankly a key dif-

ferentiator in the SchemaLogic solution, because it allows for us to control the process of how a group, an individual, a community or even an enterprise strives toward the common version of definitions, knowledge, know-how and corporate memory.

The interface for collaborative communication includes a number of features. Users of this interface may require training to help them take full advantage of the SchemaLogic system. The users who do not tap the collaborative power of the system can enter an index terms. The SchemaLogic system administrator or a trained specialist is able to resolve conflicts if term clashes arise.



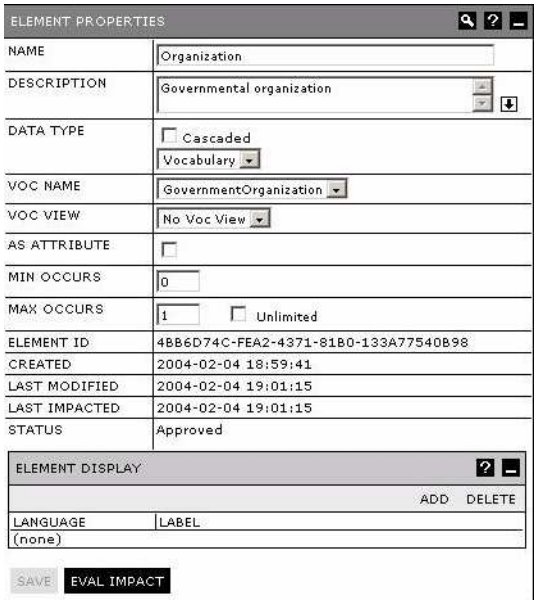

SchemaLogic provides a graphical interface for metadata management tasks. The system offers four “interfaces”: a workshop for setting up taxonomies, a catalog interface to browse knowledge bases, an ActiveTerms interface for editing knowledge bases, and customized interfaces tailored to the needs of the client.

## *Workshop for Building Data Models*

Skilled taxonomists need to build a schema or edit an existing schema. The available terms (generated manually by subject matter experts or available in an existing controlled term list) can be modeled, reviewed, moved, promoted, and demoted to create a hierarchy, terms, related terms, and other components required for a suitable controlled term list. If an ANSI-standard controlled term list is required, the Workshop facilitates the construction of this type of vocabulary.

The Workshop is a user interface designed to allow users who need to define and govern the data models. Workshop is a desktop graphical application which includes functions required to import, model, rationalize, and manage the synchronization of metadata models, schemas, and business semantics.

The Workshop provides a user-friendly way to interact with the object-oriented data modeling environment used by SchemaLogic. A user can manipulate relationship types, extension property fields, import templates, and data views.




For each term, the SchemaLogic system provides the system administrator or subject matter expert with a properties form. The form controls the properties of a specific term.

## Extending and Integrating SchemaLogic

**“"We have seen significant growth in our customer base throughout the past year," said Jeff Dirks, president and CEO of SchemaLogic. "Our customers operate information intensive businesses that rely on SchemaLogic to help them optimize information access and distribution of content. These resources will help us extend our marketing and sales outreach globally to meet specific growth goals in key markets. "—** Source is the Verity news release at http://goo.gl/Q7ilR2

SchemaLogic is a Microsoft-centric system. The system administrator can extend and integrate the SchemaLogic Server into an organization’s enterprise software ecosystem. A number of tools and services are available to make it possible to accomplish these tasks. Scripts and code are required to enable SchemaLogic Server to manage, control, and support indexing that takes advantage of the centralized controlled vocabularies in the SchemaLogic system.

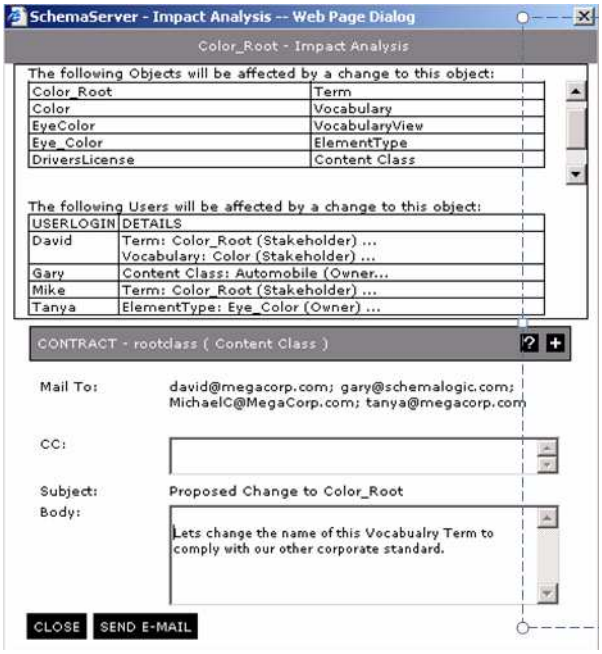
### *Software Development Kit*

The SDK (software development kit) provides developers with a set of service utilities with documentation and sample code. The SDK allows customers to integrate the semantic and structural models with other enterprise systems. The SDK allows licensees to create import filters or “adapters” to manipulate file types not supported by the system’s built in filters. Also SchemaLogic offers professional services to customers working with the SchemaLogic Enterprise Suite SDK.

### *SOAP API*

The SOAP application programmers interface is accessible from Java, Dot Net, JavaScript, and other SOA-compatible languages.

The API can make calls to the principal features and services of the SchemaLogic components.



# ArnoldIT Opinion

SchemaLogic offers a server that enforces controlled term use in an enterprise. The company has expanded the jargon about this important, but traditional approach to indexing, to obfuscate the bedrock function. The company confuses users with its “white papers” that imply semantic operations, search, text mining, and other technologies not directly germane to managing controlled vocabularies.

SchemaLogic is a system that uses look up tables, XML, and controlled vocabularies combined with structural interoperability via a thesaurus and XML schema.

Top down rigid taxonomies are really all they provide with some very lame change control and there are much better and more affordable solutions.

Table 3: SchemaLogic Checklist

| Attribute | Siderean Asserts | ArnoldIT Comment |
|---|---|---|
| 1 Platform | Microsoft-centric system | SchemaLogic is a Microsoft gold-certified partner |
| 2 Keyword search | The controlled term list can be searched | |
| 3 Text mining | No | Thi9rd-party text mining systems can process the metadata in the SchemaLogic repository |
| 4 Automated indexing | Applies terms from the controlled term list | SchemaLogic is a support system to third-party automatic indexing systems |
| 5 Personalization | Yes as a support utility | Metadata can be used to personalize third-party systems |
| 6 Workflow | Licensee configures the system to enforce metadata controls by application. | The system can automatically pull and push metadata from third-party system. |
| 7 Interface | Graphical | Scripting and coding may be required for some administrative tasks |
| 8 Hosted service | No | |
| 9 Administrative interface and tools | Graphical interfaces are provided | |
| 10 Application programming interface | Yes | |
| 11 Professional services | Yes | |
| 12 Security | Uses operating system security | |
| 13 Connectors | IBM OminiFind, IBM Content Manager, EMC Documentum, and Microsoft Office, Microsoft SharePoint, Oracle | Custom connectors can be integrated with the system via the application programming interface |

| Attribute | Siderean Asserts | ArnoldIT Comment |
|---|---|---|
| 14 Support for structured data | Structured content must be presented to the system in a format that it can process. | |
| 15 Relevance ranking | Yes | Hit boosting supported |
| 16 Video | Metadata can be controlled on a per binary object basis | |
| 17 Federated search | Yes | The metadata in the repository can be searched |
| 18 Fielded search | Yes | Search function is limited to metadata in the repository |
| 19 Content crawler | Yes | Crawls for metadata |
| 20 Price | Begins at $150,000 | |

## Anticipated Benefits

For an organization with different systems using different indexing terms, SchemaLogic can help standardize the indexing. The system prevents use of index terms that are not in the controlled vocabulary. Consistent indexing allows users of enterprise systems to locate needed information more easily. The SchemaLogic system, when properly implemented and maintained, can reduce the time required to assemble relevant information to support a work process.

SchemaLogic makes it easier to index content. A user does not have to come up with a suitable index term.

When implemented in a measured way, the SchemaLogic system can demonstrate that better indexing delivers a financial payback.

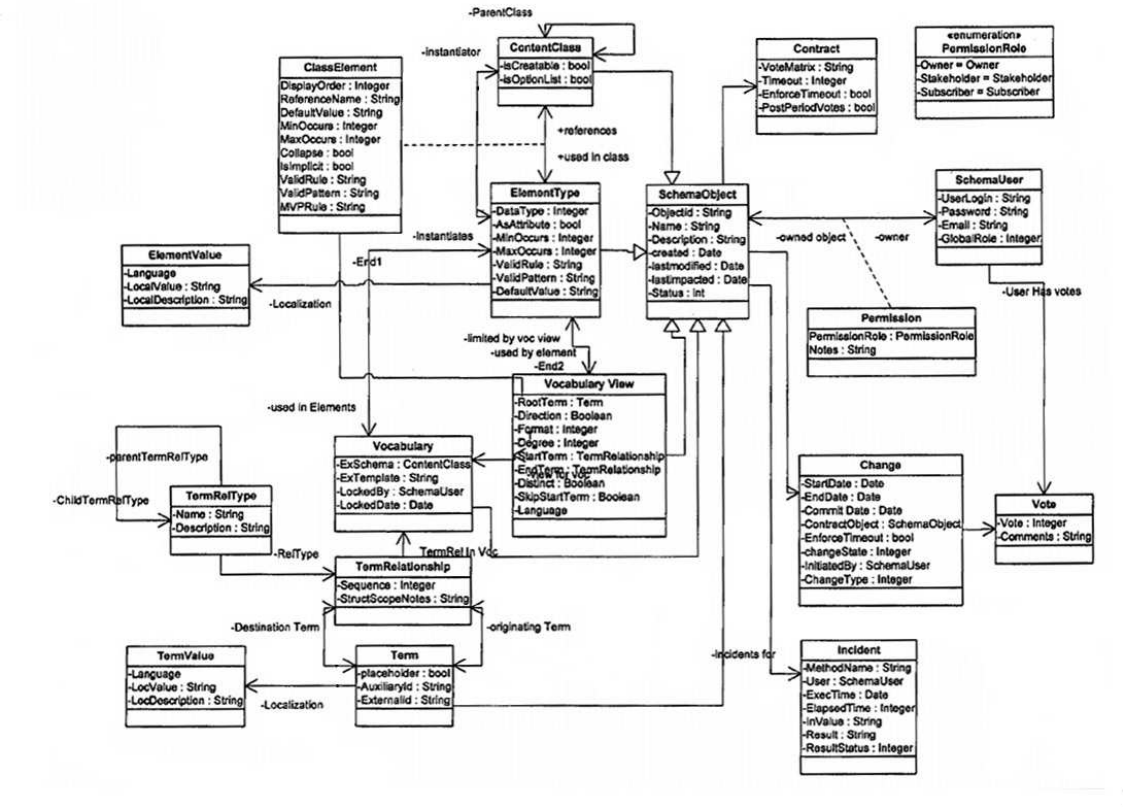
Other upsides to the SchemaLogic system include:

- Puts in place a consistent indexing foundation to minimize index drift
- Graphical interface for mapping terminology and semantics during set up and maintenance of the controlled terms and their relationships.
- Deploy a consistent, cross-system metadata repository
- Enforce tagging and nomenclature standards
- Offer a collaborative process to normalize metadata.

The workflow and integrator components of the SchemaLogic solution provides processes and technology that enables IT workers to synchronize metadata across different enterprise systems.

## Possible Drawbacks

The SchemaLogic products make sense when an organization recognizes the need for strict indexing methods. The organization's management must make a conceptual and financial commitment to implement managed controlled term indexing across disparate system. Complexity is one distinguishing feature of the SchemaLogic approach. While technically sound, those unfamiliar with taxonomies and controlled term lists may struggle to understand what the SchemaLogic system requires to function.




This SchemaLogic diagram displays the relationship among the elements of items in a controlled term list. The simplicity of a list of terms is the tip of the SchemaLogic iceberg. Numerous, detailed properties must be set up for each item in the controlled term list. Developing this type of controlled term list requires appropriate resources in the form of trained staff and funding. Time and cost are possible drawbacks for some potential licensees of the SchemaLogic system.

For an organization that makes use of an automated indexing system, the system is likely to incur significant on-going costs. Humans with indexing expertise must intervene and edit the controlled term lists and term relationships.

Other downsides include:

- Users have to be motivated to assign index terms to their documents if an automated prompting mechanism is not in place
- Subject matter experts are required to maintain the indexes
- A SchemaLogic installation requires dedicated hardware and a careful configuration and deployment process. A short cut can create a trouble-shooting headache for overworked information technology professionals.
- Your team, SchemaLogic, or a third-party integrator will have to dot the "i's" and cross the "t's" to ensure that you get the functionality you want with a minimum of custom scripting
- Custom scripting is required to integrate SchemaLogic into enterprise systems not directly supported by the SchemaLogic server's connectors.

## Net Net

Consistent metadata certainly reduces the contradictory or misleading results returned from certain search and text mining processes. It is difficult to get an argument from a librarian or a person trained in information retrieval over whether a consistent set of metadata is essential, but some managers may wonder if it is possible to bring order to enterprise metadata due to the pace of change in organizations today.

My view is that companies providing search and content-processing related systems are somewhat easier sales. Once an organization realizes that indexing has to be controlled, then an opportunity exists for a company like SchemaLogic to make a sale. The company's technology, therefore, is good, but the outlook for substantial growth is cloudy.

SchemaLogic's systems can organize and make available in one enterprise application the word lists, taxonomies, and metadata needed to improve overall accuracy of information systems in an organization. The idea is that disparate systems tap into the semantic and knowledge repository generated and maintained by SchemaLogic's system, thus reducing ambiguity and inconsistency in indexing and tagging processes.

It is difficult to disagree with SchemaLogic's premise that indexing and taxonomies must be consistent. However, not every organization is likely to recognize the need for a specialized tool for metadata management. As organizations understand the limitations of purely automatic indexing and classification systems, the value of SchemaLogic's tools is easier to grasp. Use of grassroots indexing in Wikis, folksonomies, and tag clouds introduces some retrieval problems and data integrity issues for text mining. SchemaLogic's system can normalize the metadata, killing two

birds with one stone: search gets “better” and text-mining algorithms can process more consistent metadata.

With SchemaLogic you can establish, control, and evolve a company-wide information model or enterprise taxonomy. SchemaLogic’s approach combines in one system a framework with a fairly static set of classifications, codifications, and information relationships and a mechanism to allow people to really collaborate on the evolution of those semantics.SchemaLogic brings a discipline to the uncontrolled metatags found in a folksonomy-type or Delicious.com style of system.

The system seems to be well-suited to perform metadata change management tool with secondary value as a data integration system.

Stephen E Arnold
Minor edits to a rough draft on November 26, 2013