

VENDOR PROFILE

SVP.21: H5 — Combining Technology and Process Expertise to Facilitate Information Retrieval for Information Retention and eDiscovery

Vivian Tero

IDC OPINION

H5 is unique as it offers a solution which combines the expertise from a diverse set of disciplines (to understand patterns in social relationships and human behavior to develop their hypotheses about a case) with its technology expertise in search and information retrieval to automate the search, retrieval, and analysis across large volumes of ESI in connection with litigation, investigations, and proactive information management. It is one of a handful of solution providers that addresses the eDiscovery challenge using this approach. In theory, combining automation with human knowledge and understanding the nuances of each matter facilitates a smarter approach to conducting legal reviews across large data sets. These capabilities are particularly valuable under very aggressive timelines and complex matters. The results are potentially lower document review and production costs as well as enhanced legal strategies. Fraud investigations in the finance, housing, and energy trading sectors, product liability matters in pharmaceuticals and life sciences sectors, intellectual property matters in information technology and manufacturing, and environmental safety matters in energy and utilities will continue to underpin demand for H5's solution offerings. In addition, the adoption of the Sedona Conference recommendations on best practices for search and information retrieval in eDiscovery, combined with increased awareness of the limitations of keyword/Boolean approaches, developments in case law, and the implications of the Section 502 of the Federal Rules of Evidence will compel many corporations to rethink their current approaches to search and information retrieval protocols for retention and eDiscovery. IDC also recommends that H5 take the following actions:

- Aggressively pursue opportunities in various regulatory and government agencies (if it hasn't already done so).
- Highlight that although it can automate large-scale review and information retrieval, it is not a "black box," and that the H5 domain and subject matter experts can function as 30(b)(6) witnesses, when needed.
- Expand the footprint of its "Legacy Information Management" and "Proactive Information Management" service offerings by increasing its touch points with the IT groups of its target clients (especially legal technology operations); and by expanding its current channel ecosystem and strategic partners — initially focusing on the major content management and archiving vendors.

IN THIS VENDOR PROFILE

This IDC Vendor Profile evaluates H5. The document explores the company's history, product positioning, and competitive posture. This Vendor Profile is part of the Small Vendor Profile (SVP) Series, a special IDC initiative of IDC's storage hardware, storage software, semiconductor, and compliance infrastructure groups. The goal of the SVP Series is to profile smaller "have you ever heard of" companies that have unique but compelling positions within these IDC services. Every client of at least one IDC service within one of the IDC groups listed will receive each and every vendor profile, regardless of which analyst authors the profile. The SVP Series runs from October 2008 through September 2009.

SITUATION OVERVIEW

Company Overview

H5, headquartered in San Francisco, was founded in 1999 and has over 250 employees (as of publication date of this document). Privately held H5 includes Draper Fisher Jurvetson, IVP, and Walden VC as its investors. Annual revenue is not publicly disclosed but H5 claims it is profitable and has grown 400% year on year since 2003. Engagements with H5 range from a few thousand dollars for consulting work to \$10 million to \$15 million for multiyear, enterprisewide information management services. Today, the majority of H5's engagements fall between \$200,000 and \$2 million, and last between two and eight months. IDC estimates annual revenue in the \$35 million to \$40 million range in 2008.

H5 started out as a Web search tool in 1999 but switched to legal search and retrieval services in 2002. The name H5 is attributed to John Harrison, the 18th century craftsman who solved the longitude problem by inventing the H4 chronometer. The company combines advanced proprietary technologies with professional expertise in linguistics, statistics, computer science, law, information technology, process engineering, and ediscovery to solve a corporation's litigation, investigations, and legal information management needs.

The following individuals make up the H5 senior management team:

- ☒ Nicholas Economou, Chairman and CEO
- ☒ Jeff Kangas, Executive Managing Director, head of Professional Services Group
- ☒ Eoin Beirne, Executive Managing Director, head of Operations and Technology Group
- ☒ Julia Brickell, Esq., Executive Managing Director and General Counsel
- ☒ Steve Whitehurst, Executive Managing Director, head of Global Practice Management
- ☒ Robert S. Bauer, PhD, Chief Technology Officer

H5 also has strategic advisory and scientific advisory boards, which are made up of recognized leaders in the legal, academic, business, computer science, and artificial intelligence fields. The bios of H5 senior management, strategic advisory, and scientific advisory board members are available on the corporate Web site www.h5.com.

Company Strategy

IDC sized the digital universe at 281 exabytes in 2007 and forecasts this to reach 1.8 zettabytes in 2011 with the explosion of digital cameras, sensor-based tools, datacenters supporting cloud computing, and social networks (see *The Diverse and Exploding Digital Universe*, IDC white paper, March 2008). The study also concludes that by 2010, 70% of the digital universe will be created by individuals, although corporations will have responsibility for the retention, disposition, and security of 85% of the information. The confluence of the following developments underscores the need among corporations to become smarter about their information retention, search, and retrieval activities:

- ☒ Continued explosion in digital content or electronically stored information (ESI)
- ☒ Introduction of new storage media and new content applications in enterprise networks
- ☒ Distributed, heterogeneous, and geographically dispersed information infrastructures
- ☒ Short timelines and emphasis on consistent policies and technical protocols for managing ESI under the 2006 amendments to the Federal Rules on Civil Procedure for eDiscovery

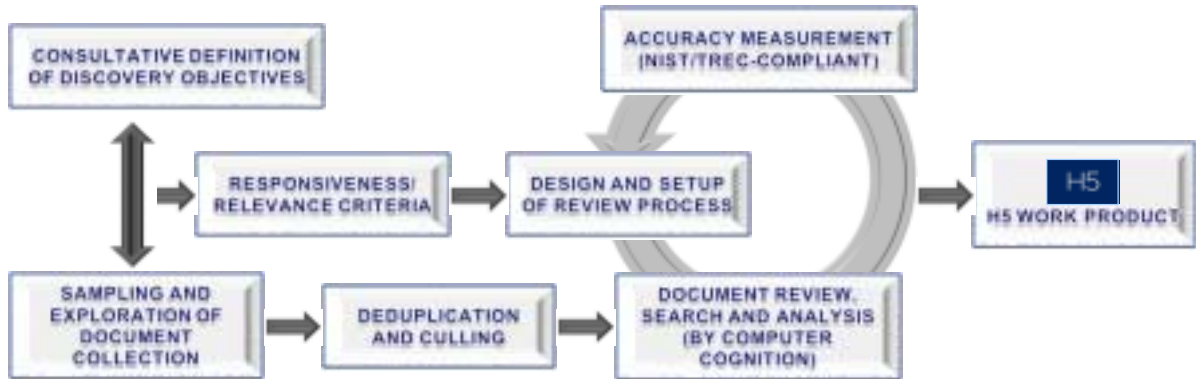
The H5 services are positioned to address the eDiscovery, litigation, and proactive information management strategy requirements of corporations that are involved in complex matters involving stores of ESI. According to H5, beyond searching for and retrieving ESI that is potentially relevant to a matter, the company discovers the narrative that would enable litigants to enhance their legal strategy.

H5 Solutions and Process Framework

H5 combines advisory and implementation services with its own proprietary advanced search and information retrieval technologies to assist clients in identifying, assessing, and managing ESI for litigation, internal and government investigations, regulatory reviews, and proactive information management for compliance and litigation readiness (Figure 1 illustrates the H5 process).

FIGURE 1

H5 Engagement Framework



Source: IDC, 2009

H5's engagements range in scope and size. While a typical engagement might involve a full responsive review, including review for privilege, trade secret, confidentiality, and/or issue coding, H5 also offers a range of advisory and litigation support services. Advisory services include discovery support, expert information retrieval services, and business performance improvement. Traditional litigation support services of H5 offers include collection, culling, processing, and document repository and production services.

What makes H5 unique is that it combines the expertise from a diverse set of disciplines to understand patterns in social relationships and human behavior to develop its hypotheses about a case. These experts have backgrounds in computational linguistics, machine learning, natural language processing, law, linguistics, statistics, IT, process engineering, eDiscovery (as traditionally defined — data extraction, deduplication, etc.), subject matter experts, and project management. H5 computer scientists then codify these theories into a set of algorithms to automate the search, retrieval, analysis, and management of large volumes of ESI. In theory, combining automation with human knowledge and understanding about the nuances of each matter facilitates a smarter approach to conducting legal reviews of very large data sets. These capabilities are particularly valuable under very aggressive timelines and complex matters. The results are lower document review and production costs as well as enhanced legal strategies. Fortune 50 companies in the technology, telecom, financial services, energy, and consumer products have employed H5 to work with them in intellectual property, product liability, securities, antitrust, and environmental liability matters. (H5 confidentiality agreements prevent senior management from publicly disclosing the names of their clients.)

In a typical engagement, H5 assembles a team of experts, which includes linguists, statisticians, lawyers, computer scientists, subject matter experts, and natural language experts. The experts deployed are based on the needs of the clients. A typical engagement would involve four or more different expert competencies. The H5 team would work closely with the client's legal counsel and eDiscovery project teams

to determine the relevance criteria. During this phase, H5 may also provide a benchmark analysis that includes one or more of the following:

- ☒ Assessment of the overall effectiveness of the clients' existing capabilities, processes, and technologies
- ☒ Evaluation of the identification and quantification of the impact of gaps between current and desired states
- ☒ Development of a process and technology roadmap for implementing possible improvements
- ☒ Implementation services (upon client's request)
- ☒ Quality control analyses of the reasonableness and soundness of the steps taken in designing or executing new or improved processes

Once the relevance criteria is defined, H5 designs and tests the protocols for search, analysis, and review. The H5 analytics and information retrieval algorithms are then programmed to find the relevant documents and automate the assessment process. The results are then tested for accuracy using NIST/TREC-compliant metrics of precision and recall, and compared against baseline metrics from the initial benchmark analysis.

H5 has an all-in fee structure that is derived from benchmark analyses to ensure that clients realize their budgetary objectives and cost savings as illustrated in the cost benchmarks.

H5 also offers a financial benchmark service that clients can use to estimate the cost of eDiscovery under various search and retrieval approaches. According to the vendor, after the clients conducted the cost benchmark, 9 out of 10 retained H5, because in doing so, they realized cost savings of 20–50% over their best alternative. Predicting the cost of eDiscovery can be very challenging when large volumes of ESI are involved, and when the merits and narrative of the case are murky. The intelligence and visibility that H5 is able to provide can be very valuable to clients involved in high-stakes litigation.

H5 Services

H5 services can be grouped into three major categories:

- ☒ **Legacy Information Management.** Legacy Information Management applies the H5 solution framework to automate the assessment of information across legacy backup tapes, and limit to a minimum the number of tapes subject to a restoration. The services include data landscape consulting; identification, inventory, and collection of physical media, sampling strategy, and execution; content cataloging; segregation of legacy information that is accessible or likely to be relevant; deduplication, culling, and filtering; media restoration into native formats; extraction and processing; and document review and categorization based on client-defined criteria. In addition, H5 would work with the client to develop a reasonable and repeatable process to segregate and dispose of

legacy information that is inaccessible or unlikely to be relevant. H5 also offers IT consulting and data management, including hosting, data transfer and migration, and workflow integration with archives or litigation repositories. Benchmark and financial analysis is also offered to gauge the reasonableness and soundness of the steps taken in the process. H5 competes with RenewData and Iron Mountain, both of which provide tape auditing, restoration, migration, and disposition/data destruction services.

☒ **Proactive Information Management.** H5 works with the client to develop and enforce legally defensible preservation protocols, and employ H5 process and technologies for eDiscovery. Services include a review of the client's data collections and summarize current state (and gaps) of each data source (including live, archival, and legacy data); legal hold and preservation process evaluation and redesign; document review (for privilege, confidentiality, and other custom attributes). H5 also provides IT consulting and data management services. This includes collection, data extraction and processing, hosting, data landscape assessments, data transfer and migration, and workflow integration with litigation repositories. H5 also designs and executes strategies to capture and manage potentially relevant data on backup tapes (legacy information management services). H5 also advises clients on how to design a stable set of categories for relevance to legal holds that enables clients to consolidate all legal holds, no matter how numerous or complex. According to the vendor, the legal hold categorization service is a real competitive differentiator and is often cited by clients as a primary reason for retaining H5. Competition in this segment is highly fragmented as risk management and litigation consulting firms, such as FTI Consulting, KPMG, PwC, Navigant, Huron, Fios, and the like, offer competitive services.

☒ **Litigation and Investigation Services.** This includes advisory and implementation services for identifying, assessing, and managing ESI for litigation, investigations (internal and government), and regulatory reviews. Services include developing culling and filtering strategies, conducting early case assessment, Rule 26(f) meet and confer consulting, document review for production, document review in preparation for deposition and trial, and IT consulting and data management. H5 also provides analytical reports that identify "hot" documents, communication patterns of interest, key potential witnesses, and gaps in production that legal teams can use to refine their document review and legal strategies. H5 also advises on the defensibility of search methods and on strategies for defending or challenging various search methods. In light of the recent case law and changes in the Federal Rules of Evidence 502, IDC opines that H5's expertise in information retrieval will provide the firm with a competitive differentiation. H5 competes with firms such as Cataphora and Digital Mandate in providing automated document review of large data sets. .

Marketing and Promotion

H5 is actively involved in various industry-sponsored and government-sponsored initiatives such as:

- ☒ **Text Retrieval Conference (TREC).** TREC is an academic/scientific initiative sponsored by the National Institute of Standards and Technology (NIST) and the U.S. Department of Defense. TREC was established in 1992 to support research within the information retrieval community. In 2006, TREC developed Legal Track, designed to evaluate the accuracy of document review and search technologies utilized in ediscovery. H5 supports and promotes information retrieval protocols used by TREC, and participated in the 2008 TREC Legal Track. The 2008 Legal Track results show that H5's document review method performed more than twice as well as the second-best entry.
- ☒ **The Sedona Conference.** The Sedona Conference is a nonprofit, 501(c)(3) research and educational institute dedicated to the advanced study of law and policy in the areas of antitrust law, complex litigation, and intellectual property rights. (For more information on the Sedona Conference and the Sedona Principles, please go to www.thesedonaconference.org.) H5 is a member of The Sedona Conference Working Group of Electronic Document Retention and Production.
- ☒ **Continuing Legal Education (CLE) programs.** H5 participates in Continuing Legal Education programs.
- ☒ **Other programs.** H5 senior executives and board members speak regularly at conferences, academic institutions, and ACC roundtables.

FUTURE OUTLOOK

The following developments will underpin demand for H5's solutions:

- ☒ In August 2007, the Sedona Working Group on Best Practices for Document Retention and Production (WG1), Search & Retrieval Sciences Special Project Team released a report, *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in eDiscovery*. The report recognizes that the volume and exponential growth of ESI, combined with the inherent ambiguity in the human language and social interactions (which impacts concepts and context), and the recognized deficiencies in simple keyword searching pose challenges to effective and legally defensible search and retrieval approaches during eDiscovery. The Sedona Conference encouraged corporate litigants and legal counsels to explore innovative ways to conduct these activities. It recommends litigants to consider employing methods from other disciplines (such as linguistics and information retrieval science). It also recommends that litigants consider utilizing alternative search and text and analytics tools to supplement their simple keyword and Boolean search techniques. The Sedona Conference also recommends that litigants agree on methods to measure and evaluate the effectiveness of the search and retrieval process.
- ☒ Recent developments in case law and in the Federal Rules of Evidence 502 may compel more corporations to reevaluate their current search, retrieval, preservation, and collection approaches. These developments further validate the recommendations of the Sedona Conference Working Committee on the use of search and information retrieval methods in eDiscovery.

- ❑ The court's opinions in *United States v. O'Keefe*, and *Equity Analytics, LLC v. Lundin* centered on the use of search terms for conducting a diligent search for potentially relevant ESI. Both cases suggested that litigants cannot simply rely on their legal counsels to determine the search protocols. It highlighted the criticality of expert testimony in defining the search protocols. (For more on this, see *United States v. O'Keefe*, 537 F. Supp. 2d 14, 24 [D.D.C. 2008], and *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331 [D.D.C. 2008].)
- ❑ In *S.E.C. v. Collins & Aikman Corp.* No. 07 Civ. 2419 (SAS), Judge Shira Scheindlin's opinion in the SEC's disfavor (for overproduction) articulates a new standard for collaborative search. Judge Scheindlin opined, "...At the very least, the SEC must engage in a good faith effort to negotiate with its adversaries and craft a search protocol designed to retrieve responsive information without incurring an unduly burdensome expense disproportionate to the size and needs of the case."
- ❑ On September 19, 2008, President Bush signed S. 2450 into law. This included the new Evidence Rule 502 to the Federal Rules of Evidence. This new rule protects against the inadvertent waiver of attorney-client privilege or the work product protection, and applies to all proceedings after the date of enactment and those pending on the date of enactment. This new regulation underscores the importance defining the search and preservation protocols as part of the litigants' efforts to conduct reasonable steps to prevent inadvertent waiver of attorney-client privilege. It also highlights the need for relevant subject matter experts and not simply rely on legal counsel to determine the Boolean keyword searches for conducting privilege reviews.

The expected increase in lawsuits and regulatory investigations resulting from the recent economic meltdown continues to open up opportunities for solutions offered by providers such as H5.

Advanced analysis and review capabilities are typically employed by corporate litigants postcollection; and postcollection document review and analysis has been the core of H5's solution. But the H5 solution could potentially enable corporate litigants to utilize these resources during the information management, identification, and preservation phases of eDiscovery. The TREC results indicate that such an approach enhances the accuracy of the search for and retrieval of potentially relevant ESI, thereby mitigating the risks and document review costs from overpreservation and overcollection. It could also potentially enable the corporate litigant and its counsel to develop their legal theories and strategies about the case very early on. Corporations that are involved in complex matters involving very large data sets would view this as a strategic tool that they could utilize to build on their legal strategy. Corporate litigants are also able to improve the precision of their forecasts on the total expected costs of eDiscovery and plan for it accordingly.

ESSENTIAL GUIDANCE

Advice for H5

Despite the impressive growth in H5's revenue in the past four years, large-scale automated document reviews have been relatively slow to gain traction. Corporations and law firms that have to review and manage multiple terabytes of ESI and are involved in complex high-stakes matters are the ones most likely to have the appetite for H5's services. In the near term (6–18 months), IDC expects that this will continue to be the trend among corporate litigants. However, media reports suggest that U.S. regulatory government agencies such as the SEC, FBI, and FTC are overwhelmed with the volume of fraud investigations. H5 is positioned to take advantage of these opportunities (if it hasn't already done so). Large-scale automated document reviews was the core of H5's business in 2003 to 2007, but the developments in eDiscovery also open up several avenues where H5's core expertise in information retrieval will be valuable.

The supplier landscape for eDiscovery products and services remain highly fragmented in spite of the continued market consolidation. H5 has, so far, differentiated itself with its hybrid "human and process experts + technology" approach in addressing the eDiscovery and information retrieval needs of clients for complex litigation, investigations, and proactive information management. It should continue to make investments in these areas. H5 should also highlight that although it can automate large-scale review and information retrieval, it is not a "black box," and that the H5 domain and subject matter experts can function as 30(b)(6) witnesses, when needed. This is a misconception that H5 would need to address more aggressively.

H5 has primarily been working with outside law firms to address the document review and legal strategy needs of its clients. IDC recommends the following actions as it looks to expand the footprint of its legacy information management and preservation and litigation readiness service offerings:

- ☒ Increase its touch points with the IT groups of its target clients (especially legal technology operations) (At the end of the day, IT plays a key role in ensuring that the technologies and solutions they deploy are a good fit for the process and technical maturity of their IT infrastructure.)
- ☒ Expand its current channel ecosystem and strategic partners to include the content management and archiving vendors

LEARN MORE

Related Research

- ☒ *How Are Corporations Managing ESI for Legal Holds?* (IDC #215424, November 2008)
- ☒ *Worldwide Email Archiving Applications 2008–2012 Forecast Update and 2007 Vendor Shares* (IDC #213701, August 2008)

- ☒ *Beyond Active Archiving: Opportunities from Corporate Litigation Readiness Investments* (IDC #213823, August 2008)
- ☒ *Highlights from the 2007 Information Management for Compliance Survey* (IDC #211216, March 2008)
- ☒ *Active Archiving: Retention, Preservation, and Reuse Multiclient Study Final Report* (IDC #210315, February 2008)

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2009 IDC. Reproduction is forbidden unless authorized. All rights reserved.

Published Under Services: Compliance Infrastructure; European Storage Technologies and Strategies; Hard Disk Drive Components; Semiconductors: Overview; Storage Mechanisms: Disk; Storage Mechanisms: Optical/Removable; Storage Mechanisms: Solid State Drives; Storage Mechanisms: Tape; Storage Software; Storage Systems