

Cross-Border E-Discovery/E-Disclosure Workshop (DESI III¹)
ICAIL 2009, Barcelona
Monday, June 8, 2009

Background

Throughout the world, lawyers increasingly face the problem of how to effectively and efficiently conduct searches for relevant documents in heterogeneous haystacks of electronic data. Hundreds of millions of documents are now routinely subject to searches in a variety of litigation and investigatory contexts. In recognition of this situation, the United States recently adopted new rules governing civil litigation in federal courts. These courts have recognized “electronically stored information” (ESI) as a term of art, embracing all forms of electronic documents made subject to the civil discovery process. Under the new rules, opposing parties in federal court litigation now have an early “meet and confer” duty to discuss a range of issues, including the continued storage, preservation, and access to ESI in their respective physical and legal custodies.

Of course, litigation issues involving massive volumes of electronically stored information are not confined to the U.S.: they exist in the U.K., the E.U., and around the world, as judicial systems are developing procedures and protocols for dealing with ESI. For example, in the U.K., Part 31 of the Civil Rules governs disclosure of electronic evidence in civil cases. These “rule sets” are of growing importance, not only in common law jurisdictions but also in civil law jurisdictions, such as in Spain and many other EU countries. See, e.g., the Article 29 Data Protection Working party plenary session held in Brussels in October 2008, where e-discovery/e-disclosure issues were discussed. An added layer of complexity is often introduced by cases involving multiple foreign jurisdictions, as these cases may require translation and cross-language search of foreign language text. In addition, attorneys need to understand the restrictions and requirements of how searches for information can and must be carried out under foreign law.

As one part of the discussion, courts can be expected to require that parties make good faith efforts to collaborate on the selection of search protocols and parameters. Related to this, massive collections of ESI present significant challenges to litigators and investigators tasked with identifying and making sense of the relevant information they may contain. However, most litigators find themselves ill-equipped to evaluate state-of-the-art legal tech-sector claims as to what search retrieval methods, tools, and techniques should be used. But there has been some limited recognition by the U.S. judiciary that a world of information retrieval and AI techniques exists as a possible aid to the legal profession. See, e.g., *Victor Stanley v. Creative Pipe* (U.S. District Court, Maryland, 2008): in holding that a “privilege log” of documents was inadequate in having been constructed on the basis of a faulty keyword search, the court discusses a variety of other IR techniques, including Bayesian classifiers, probabilistic search models, fuzzy search models, clustering techniques, and concept and categorization tools, as being of possible interest to lawyers; in *United States v. O’Keefe* (U.S. District Court, District of Columbia, 2008), the court noted that problems of keyword searching involve a complex interplay between linguistics, computer science and the law).

¹ An equivalent e-discovery workshop held as part of ICAIL 2007 in Palo Alto, CA was entitled “Workshop on Supporting Search and Sensemaking For Electronically Stored Information in Discovery Proceedings,” which in shorthand became the first “DESI” workshop.

Two initiatives have emerged over the past few years in the U.S. that begin to address these challenges. The first has been the work of The Sedona Conference® working group on electronic document retention and production, composed principally of legal professionals with experience in civil discovery involving ESI. The group includes both lawyers and so-called “e-discovery” firms. In August 2007, The Sedona Conference® released a worldwide public draft of a Best Practices Commentary on the Use of Search and Information Retrieval in E-Discovery. The second is a “legal track” sponsored by the National Institute of Standards and Technology (NIST) as part of its annual Text Retrieval Conference (TREC), consisting of a multi-year collaborative information retrieval research project focused on information retrieval for e-discovery applications in which both academics and corporations can participate. Results from the third year of the TREC legal track will be preliminarily available in late 2008, and will be published formally in the Spring of 2009 – with a fourth year of the legal track approved for 2009.

Additionally, two similar, successful international workshops on e-discovery have recently been held. As indicated above, the first “DESI” Workshop was held in conjunction with ICAIL 2007. It brought a wide array of individuals together to foster engagement between e-discovery practitioners and a broad range of research communities who can contribute to the development of new technologies to support the e-discovery process. A second workshop (“DESI II”) was held at University College London in June 2008. The latter broadened the scope of this discussion to include comparisons of requirements between different national settings and different legal contexts, and to foster the greater involvement of particularly U.K. researchers in these efforts. A cross-border e-discovery/e-disclosure workshop at ICAIL 2009 would provide a further platform for discussion of search issues in an E.U. and international context.

Organization of the Workshop.

The full-day workshop will be organized in five parts:

Part I will begin the workshop with a brief overview of recent developments, including the results from the first DESI workshop, results from the second year of the TREC legal track, and recent products of the Sedona Conference®. This may be followed by a discussant who will react to the points that have been made in a way that highlights differences in the European context, as well as between civil law and common law type jurisdictions. An open discussion period regarding goals for the day will then conclude this first part. The goal of Part I will be to help to shape and stimulate the discussion that will ensue throughout the day.

Part II, the second session before lunch, will consist of contributed research presentations. We plan to solicit participation from e-discovery firms as well as researchers in information retrieval and natural language processing, AI and Law, human language technology and human-computer interaction, text mining and text classification, digital forensics and archival science, information studies and legal sensemaking on a global scale. We will solicit both research contributions and position papers. Since the workshop will be part of the International Conference on AI and Law, the Call for Participation will invite presenters to discuss the question of how AI techniques can help address the challenges of e-discovery. We hope to invite some of the presenters to expand their papers and submit them to a planned Special Issue of the *Journal of Artificial Intelligence and Law* devoted to the topic of e-discovery.

Part III will be a set of moderated “breakout sessions” over lunch in which participants will brainstorm ideas for a practical, contemporary research agenda.

Part IV, immediately following lunch, will consist of invited presentations on specific issues that serve to round out the set of topics covered in part II. We expect to invite an even balance between practitioners and researchers, and to achieve the best possible balance across national settings (including non-English settings) for the practitioners and across research communities for the researchers.

Part V will consist of a panel discussion, with one panelist for each lunch table topic. The panel discussion will include brief presentations by each panelist, followed by a facilitated open forum addressing questions such as:

- What research questions should be explored that are not presently being addressed?
- How can we help to inform the practice of legal professionals, and to what extent does that task vary by jurisdiction?
- Who, beyond those already in the room, do we need to engage with to address the challenges that we have identified?

Of course, the actual questions that we address will emerge from our discussions over the course of the day. We will create both a public Web page and an open mailing list to disseminate the results of the workshop. While optimally the workshop should be kept of modest size, we expect that the diverse range of research issues and the large number of research communities involved, not to mention the amount of commercial interest in this topic, may result in many more interested participants than that. We will therefore give first preference to those who are presenting, and then accept additional participants up to whatever capacity the chosen facility can accommodate.

We are also in discussions with The Sedona Conference® as to whether they would wish to co-locate a meeting of their International Working Group (Working Group 6) in Barcelona to coincide with the ICAIL conference, so as to provide maximum opportunities and incentive for Sedona members to participate in the workshop.

Additional Information

Time and Place: The Workshop will be held as a pre-conference workshop at ICAIL 2009. It is anticipated that the Workshop will begin at approximately 08:30 and end at 16:30.

Registration: **[to be filled in as worked out]**

Workshop Organizing Committee

Kevin D. Ashley, Ph.D.
Professor of Law and Intelligent Systems
University of Pittsburgh School of Law
Pittsburgh, Pennsylvania
<http://www.lrdc.pitt.edu/Ashley/>

Dr. Kevin Ashley holds interdisciplinary appointments as a faculty member of the Graduate Program in Intelligent Systems at the University of Pittsburgh, a Senior Scientist at the Learning Research and Development Center, a Professor of Law, and Adjunct Professor of Computer Science. His goals are to contribute to Artificial Intelligence (AI) research on case-based and analogical reasoning, argumentation and explanation and to develop instructional systems for students and professionals in case-based domains such as law and ethics. He received a B.A. in philosophy (magna cum laude) from Princeton University in 1973, J.D. (cum laude) from Harvard Law School in 1976, and Ph.D. in computer science in 1988 from the University of Massachusetts where he held an IBM Graduate Research Fellowship. For his Ph.D. he developed an AI CBR system, HYPO, which reasons by analogy to past legal cases, makes arguments about legal fact situations and poses hypothetical cases. MIT Press / Bradford Books published his book based on his dissertation entitled *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*. In April, 1990, the National Science Foundation selected Professor Ashley as a Presidential Young Investigator, and in 2002 he was selected as a Fellow of the American Association of Artificial Intelligence. From June, 1988 through July, 1989, he was a Visiting Scientist at the Thomas J. Watson Research Center, Yorktown Heights, New York.

Jason R. Baron
Director of Litigation
National Archives and Records Administration
8601 Adelphi Road, Suite 3110, College Park, MD 20740
jason.baron@nara.gov, tel. +1-301-837-1499

Jason Baron serves as Director of Litigation for the U.S. National Archives and Records Administration. Previously, Mr. Baron held successive positions as trial attorney and senior counsel in the U.S. Justice Department in Washington, D.C., where he represented the interests of the U.S. government in a variety of complex lawsuits involving access to governmental information, including acting as lead counsel in two cases involving the preservation of White House email. He currently represents NARA on The Sedona Conference® Working Group on Electronic Records Retention and Production, where he serves on the Working Group Steering Committee and is Editor-in-Chief of The Sedona Conference® Best Practices Commentary on the Use of Search and Information Retrieval in E-Discovery. For the past three years he has served as a founding co-coordinator of the National Institute of Standards and Technology TREC (Text Retrieval Conference) legal track. He has been a member of InterPARES, a Visiting Scholar at the University of British Columbia, and is currently an Adjunct Professor at the University of Maryland's graduate College of Information Studies. He also presently serves on the Georgetown University Law Center Advanced E-Discovery Institute advisory board. Mr. Baron received degrees from Wesleyan University and the Boston University School of Law.

Jack G. Conrad & Marc Light
Senior Research Scientists
Research & Development
Thomson Reuters Professional
St. Paul, Minnesota 55123
jack.g.conrad@thomsonreuters.com
marc.light@thomsonreuters.com

Jack G. Conrad is a Senior Research Scientist in the Research & Development group at Thomson Reuters. His research areas fall under a broad spectrum of Information Retrieval topics. Some of these include document clustering and deduplication for knowledge management systems, resource selection in massive data environments, and document structure analysis for legal texts. Jack has researched and implemented resource discovery techniques for applications in an operational environment consisting of tens of thousands of databases and has developed and evaluated algorithms for real-time fuzzy duplicate detection in large document repositories. He has authored numerous technical papers and has several patents pending. Jack completed his graduate studies in English (Linguistics) at the University of British Columbia–Vancouver and in Computer Science (Information Retrieval) at the University of Massachusetts–Amherst. His undergraduate fields of study were Electrical Engineering and English at Marquette University in Milwaukee.

Marc Light is also a Senior Research Scientist in the Research & Development group at Thomson Reuters. For over 20 years, he has worked in the areas of natural language processing and information retrieval. Specific topics include automated lexicon induction, shallow parsing for speech-to-speech translation, question answering, passage retrieval, knowledge management for biologists, low latency named entity tagging, and named entity tagging of legal documents. Marc received his PhD and Masters in Computer Science from the University of Rochester and his undergraduate degree in Cognitive Science from MIT. Prior to starting at Thomson Reuters, he held research positions at the University of Iowa, the MITRE Corporation, the University of Stuttgart, and the University of Tuebingen.

Debra Logan
Vice President and Distinguished Analyst
Gartner, Inc
Tamesis 2
The Glanty
Egham, Surrey
UK RH176HB
debra.logan@gartner.com

Debra Logan, VP and Distinguished Analyst, has been with Gartner for nine years. Currently, she is doing research in the area of legal IT, investigating the effects of regulations and rules changes on the IT infrastructure, applications and staffing. Debra's other research interests included systems of knowledge organization, taxonomies and metadata, content, document and records management, process management and the application of enterprise information management technologies and concepts in business. During her 25-year career, she has done IT consulting and software development work all over the world with many Global 2000 companies and government organizations, including as part of her own consulting company. Debra began her career at the University of Pittsburgh's Learning Research and Development Center (LRDC), where she was a leading member of a team that developed an AI-based Intelligent Tutoring System for the US Air Force. Debra holds a Master of Science degree in Library and Information Science from the University of Pittsburgh, and a Bachelor's degree in Psychology and History from Washington and Jefferson College.

Additional References

Ashley, Kevin D., "Can AI & Law Contribute to Managing Electronically Stored Information in Discovery Proceedings? Some Points of Tangency," paper presented at DESI Workshop, Workshop on Supporting Search and Sensemaking For Electronically Stored Information in Discovery Proceedings, Eleventh International Conference on Artificial Intelligence and Law, Palo Alto, June 4, 2007, <http://www.umiacs.umd.edu/~oard/desi-ws/>

Baron, Jason R. "Toward A New Jurisprudence of Information Retrieval: What Constitutes A 'Reasonable' Search for Digital Evidence When Using Keywords?," *Digital Evidence Journal*, 2008 (U.K.) (forthcoming)

Baron, Jason R., "Toward A Federal Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery," 6 *Sedona Conference Journal* 237-246 (2005) (available on Westlaw, Lexis)

Baron, Jason R., *The TREC Legal Track: Origins and Reflections on the First Year*, 8 *Sedona Conference Journal* 251-259 (2007) (available on Westlaw, Lexis)

Collaborative Expedition Workshop #45, *Advancing Information Sharing, Access, Discovery and Assimilation of Diverse Digital Collections Governed by Heterogeneous Sensitivities*, held Nov. 8, 2005, http://colab.cim3.net/cgi-bin/wiki.pl?AdvancingInformationSharing_DiverseDigitalCollections_HeterogeneousSensitivities_11_08_05

Conrad, Jack G., "E-Discovery Revisited: A Broader Perspective for AI Researchers," paper presented at DESI Workshop, Workshop on Supporting Search and Sensemaking For Electronically Stored Information in Discovery Proceedings Eleventh International Conference on Artificial Intelligence and Law, Palo Alto, June 4, 2007, <http://www.umiacs.umd.edu/~oard/desi-ws/>

DESI Workshop, Workshop on Supporting Search and Sensemaking For Electronically Stored Information in Discovery Proceedings Eleventh International Conference on Artificial Intelligence and Law, Palo Alto, June 4, 2007, <http://www.umiacs.umd.edu/~oard/desi-ws/>

DESI II Workshop, Second International Workshop Supporting Search and Sensemaking for Electronically Stored Information in Discovery Proceedings, University College London, June 25, 2008, <http://www.cs.ucl.ac.uk/staff/S.Attfield/desi/>

Digicel (St. Lucia) Limited v. Cable & Wireless PLC (England and Wales High Court (Chancery Division), [2008] EWHC 2522 (Ch) (23 October 2008).

NIST TREC Legal Track web page, <http://trec-legal.umiacs.umd.edu> (containing TREC 2006 and TREC 2007 overview papers, with TREC 2008 conference proceedings forthcoming).

Paul, George L. and J.R. Baron, "Information Inflation: Can The Legal System Cope?," 13 *Richmond Journal of Law and Technology* (2007), <http://law.richmond.edu/jolt/v13i2/article10.pdf>.

The Sedona Conference, *The Sedona Best Practices Commentary on the Use of Search and Information Retrieval in E-Discovery* (2007 Public Draft),,see <http://www.thesedonaconference.org/content/miscFiles/publications.html>

United States v. O'Keefe, 537 F. Supp. 2d 14 (D.D.C. 2008)

Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251 (D. Md. 2008)