Call for Submissions
Seventh DESI Workshop:
Using Advanced Data Analysis in eDiscovery & Related Disciplines To Identify and Protect Sensitive Information in Large Collections
London, U.K.
Monday June 12, 2017

http://www.umiacs.umd.edu/~oard/desi7/

A Pre-Conference Workshop held in conjunction with the
2017 International Conference on Artificial Intelligence and Law

The DESI VII workshop will provide a platform for discussion of best practices and innovations in the use of advanced search technology, text classification, language processing, data organization, visualization and related techniques for the purposes of accessing and managing electronically stored information. One focus of the DESI VII workshop will be on emerging protocols and novel techniques for identifying and protecting sensitive information in large collections. The workshop will also welcome contributions on other topics that are within the workshop's broader scope. We expect the refined focus on protecting sensitive content this year to be directly relevant to at least four application contexts:

- eDiscovery in complex litigation
- European Union (EU) privacy policies
- Audits and internal investigations
- Public access to government records

We expect to address the following open questions:

In eDiscovery: What techniques are currently being used to classify information found in email or other data sources as privileged, confidential, or otherwise protected by law? How widespread is the use of technology for this type of information identification? How well do current technologies perform with respect to the classification of sensitive information?

In EU privacy policies: To what degree can current algorithmic techniques adequately characterize content that individuals might wish to have blocked from certain types of access in adherence with "right to be forgotten" laws? To what extent can the process of adjudicating such requests reasonably be automated? How well do algorithmic techniques perform in identifying sensitive data that may need to be blocked from cross-border transfers? To what extent can these capabilities satisfy requirements for algorithmic accountability?

In audits and investigations: What tools and techniques are available to find and protect well-defined categories of sensitive content? Examples from the US and Canada might include protected health information, student education records, customer record information, card holder data, or proprietary or confidential information (e.g., trade secrets). To what extent can taxonomies be constructed for information that is routinely the focus of internal audits to

facilitate automatic detection of those categories of information? To what extent can technical support for investigations be designed to protect sensitive content that is not material to the investigation?

In public access requests: How well can current procedures and automated techniques identify and protect personal, political, proprietary or otherwise confidential content? To what extent can automated techniques reliably detect specific types of personally identifiable information which, if released, would constitute an unwarranted invasion of privacy?

The workshop discussion will be grounded in the results of original research, such as that reported in interdisciplinary venues such as ICAIL, law reviews, technical conferences in specific disciplines (e.g., KDD, ICWSM, ACL, SIGIR), and shared task evaluations (e.g., TREC, CLEF, NTCIR).

Participation is invited from all interested parties, including those with backgrounds in:

- Archives and records management
- Artificial intelligence and law
- Cognitive science
- Content analytics
- Corpus analysis
- Computational linguistics
- Digital forensics
- eDiscovery
- Human-computer interaction
- Human language technology
- Information governance
- Information retrieval
- Knowledge management
- Legal informatics
- Legal sensemaking
- Litigation support
- Natural language processing
- Machine learning
- Privacy
- Sentiment analysis
- Text mining and classification
- Visual analytics

Submissions:

Two types of written contributions are invited:

- *Research & Operational Practice Papers*. Original papers (limited to 4 to 10 pages) describing current research results, experimental or emerging practices, or current best practices. Research and operational practice papers will be peer reviewed separately. After peer review, accepted papers will be posted on the DESI VII website. Authors of accepted operational practice or research papers will be invited to present their work either as an oral or a poster presentation. These papers are due on April 1, 2017; decisions will be returned by May 1, 2017.

- *Position papers* (limited to 1-5 pages) describing individual interests, for inclusion on the DESI VII web site and distribution to workshop participants. Submissions of this type are particularly valuable when bringing together diverse research communities. Additionally, these papers can help with our selection of discussion leaders and panelists. Position papers are not peer reviewed, but there is an editorial review to ensure that they satisfy the 5-page length limit and that they address one or more topics within the broad scope of the workshop. Position papers are requested by May 1, 2017. Participation in the workshop is open, so while prior submission of position papers is strongly encouraged it is not strictly required.

Please note that because of the workshop's focus on research interchange, we are not able to accept commercial white papers or similar corporate materials.

Submissions should be sent by email to Jack Conrad (jack.g.conrad@tr.com) with the subject line DESI VII RESEARCH/OPERATIONALPRACTICE PAPER or DESI VII POSITION PAPER. All submissions received will be acknowledged within 3 days.

Organizing Committee:
Jason R. Baron, Drinker Biddle & Reath LLP
Jack G. Conrad, Thomson Reuters
Hans Henseler, University of Applied Sciences, Leiden
Amanda K. Jones, H5
Douglas W. Oard, University of Maryland

Details on our **Program Committee** to be announced.