J18 A Novel Automated, Searchable Database for the Chemical Characterization and Comparison of Printing Inks

Tatiana Trejos, PhD, Florida International University, Chemistry Dept, 11200 SW 8th Street, University Park, OE 109, Miami, FL 33199; Peter Torrione, PhD, CoVar Technologies, 1495 Chain Bridge Road, Ste 100, McLean, VA 22101; Ruthmara Corzo, BSc, Florida International University, 11200 SW 8th Street, Miami, FL 33199; Paul Martin, PhD, CRAIC Technologies, 948 N Amelia Avenue, San Dimas, CA 91773; Anna Raeva, PhD, 11200 SW 8th Street, CP194, Miami, FL 33199; Kiran Subedi, 1191 NW 7th Court, Apt 2, Miami, FL 33136-2331; Rhett J. Williamson, 1100 SW 104th Court, Apt 302, Miami, FL 33174; Jong Yoo, PhD, 46665 Fremont Boulevard, Fremont, CA 94538; and Jose R. Almirall, PhD, Florida International University, Dept of Chemistry, University Park, Miami, FL 33199*

After attending this presentation, attendees will be informed about a new database that was created for the automated search and comparison of printing ink evidence. The database contains spectra from the following six analytical methods using standardized acquisition parameters for each method: (1) Attenuated Total Reflectance Fourier Transform Infrared (ATR/FTIR) spectroscopy; (2) Raman spectroscopy; (3) Scanning Electron Microscopy with Energy Dispersive Spectroscopy (SEM/EDS); (4) Pyrolysis Gas Chromatography/Mass Spectrometry (Py-GC/MS); (5) nanosecond Laser Ablation Inductively Coupled Plasma/Mass Spectrometry (ns-LA-ICP/MS); and, (6) Direct Analysis in Real Time Mass Spectrometry (DART®-MS). The relative discrimination and overall performance of the database search routine, including error rates (false associations and incorrect discrimination), are reported for the first time for each of the six analytical methods that make up the chemical database as well as for combinations of the data (data fusion). Attendees will gain knowledge on the relative utility and performance of these analytical methods for the examination of different types of printing inks.

This presentation will impact the forensic science community by providing a comprehensive evaluation of the performance of a searchable database that was designed and validated as a tool to augment the chemical information mined from ink evidence and to aid in document examinations and related investigations. The software will become commercially available for use and it is expected that the database will expand with the number of users, making the database increasingly useful through the continuous feedback from examiners.

The database contains two main sets of data: the reference database and the training or testing database. The reference database currently contains 4,658 data files acquired from representative samples from selected printing sources to account for some of the global variety in printers, including 319 toner, inkjet, offset, and intaglio inks. The testing database contains over 800 data files acquired from 45 duplicate control samples used to train the searching algorithms. Duplicate controls were analyzed over different days to evaluate correct associations and to assess the potential for instrumental variation, operator variation, and within-sample heterogeneity.

The developed software makes use of machine-learning algorithms for classification and comparison of unknown samples to the database collection, in particular Partial-Least Squares Discriminant Analysis (PLSDA) and K-Nearest Neighbor (KNN) spectral comparisons. The search algorithms generate similarity scores that permit the operator to significantly narrow down the possible sources of ink samples contained within the database. The user may select to use any number of methods depending on the case, type of ink, sample size, or other factors. Data fusion algorithms were also built into the database to allow for data fusion from two of any of the analytical techniques, often providing improved performance for fused data searches over searches using any technique in isolation. The relative discrimination of all the analytical methods has been determined for each ink type. LA-ICP/MS provides the best discrimination (>99%) and py-GC/MS provides the least information, regardless of ink type. The discrimination capability of SEM/EDS was relatively poor for inkjets and offsets but was found to be good for toner and intaglio samples, with discriminations of 97.2% and 98.2%, respectively. Although FTIR and Raman provided lower discrimination capabilities, they are useful for characterization of major organic components and classification or grouping of inks.

A validation study of the database search algorithms, using blind searches of duplicate controls, reveals that LA-ICP/MS and FTIR data searches result in the best performance for associating duplicate samples analyzed by different operators on different days with >90%-100% of duplicate pairs correctly associated. DART®-MS data search results in 83%-100% of the duplicate pairs correctly associated, with the other analytical techniques resulting in 33%-100% correct associations for SEM/EDS and 6%-81% for Raman, respectively. The lower correct association performance is often related with a limited number of duplicate controls available in the current testing database for particular sensors and/or particular limitations of certain techniques for some types of ink.

Although the current collection set represents a relative small snapshot of the printing inks universe, it serves as a proof of principle of its potential utility and relevance in the field. This novel chemical database now provides a comprehensive inorganic and organic characterization of different printing inks and permits data fusion of the multiple sensors.

Printing Inks, Database, Data Fusion