

data VISUALIZATION

What is it? And can you trust it?

BY GEORGE J. SOCHA

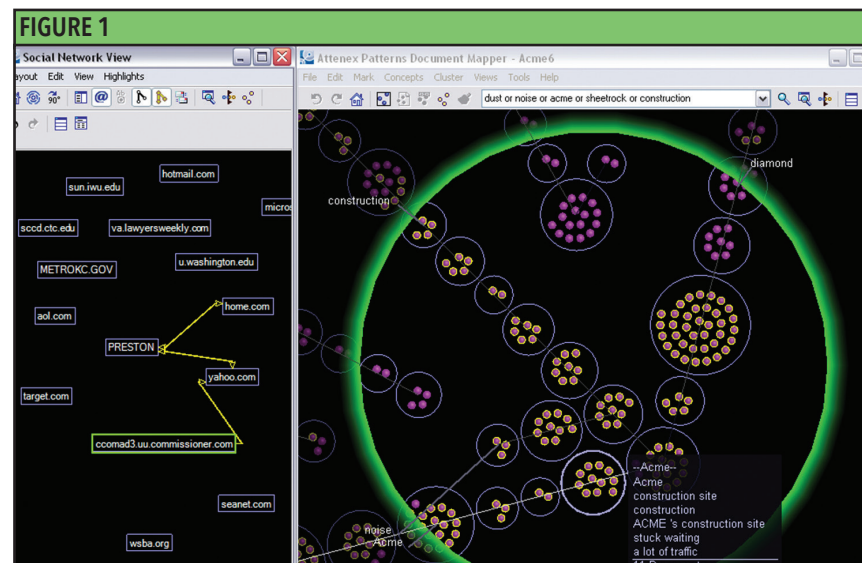
THE LEGAL TECHNOLOGY PRESS IS REPLETE WITH ARTICLES TOUTING THE ADVANTAGES OF TECHNOLOGY ASSISTED REVIEW (TAR), MOSTLY AS A WAY TO STREAMLINE THE REVIEW OF ELECTRONICALLY STORED INFORMATION (ESI) AND TO HELP REDUCE REVIEW COSTS. Recently a growing number of articles have pushed into an area of TAR both new and old: the use of data analytics, especially data visualization, across all stages of the EDRM (E-Discovery Reference Model).

According to the TechTarget website (www.techtarget.com), a popular resource for techies, data analytics “is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software.”¹ The article goes on to note that “analytics technologies

and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions and by scientists and researchers to verify or disprove scientific models, theories and hypotheses.”

Data visualization, to cite TechTarget again, “is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software.”²

Data visualization is well established in the business, scientific, and educational communities. For anyone seeking to learn more about the broader uses of data visualization, I highly recommend the books published by Edward Tufte, a statistician, artist, and professor emeritus at Yale University,



THESE VISUAL CLUSTERS CAN BE USED FOR MANY PURPOSES, BUT PERHAPS THE MOST COMMON USE IS TO HELP PUT SIMILAR DOCUMENTS INTO GROUPS TO STREAMLINE THE REVIEW PROCESS. IT WILL BE EASIER AND FASTER FOR A SINGLE REVIEWER TO LOOK AT A GROUPING OF 500 SIMILAR DOCUMENTS THAN TO REVIEW THOSE SAME DOCUMENTS INTERSPERSED AMONG 4,000 OTHER DOCUMENTS.

especially *The Visual Display of Quantitative Information*, which was first published in 1983.³

Litigators long have had access to tools that place ESI in context. These tools work with the contents of files (the text of e-mail messages and word processing documents, for example) and metadata about those files (authors, dates messages are sent or created, etc.). An early example was *Attenex Patterns*,⁴ which I first used in 2003, a tool often referred to both with praise and derision as the “Petri dish.” (See *Figure 1*.)

A variety of data visualization tools are available to help counsel and clients better understand the significance of ESI. Such tools draw attention to patterns, trends, and correlations in the ESI that help litigation departments prepare their cases. These include tools specifically built for e-discovery, such as kCura’s *Relativity*⁵ and *Brainspace Discovery*,⁶ as well as tools common in the larger world of data analytics, such as *Tableau*.⁷ These tools are very well used: According to kCura, for example, more than 150,000 people at 12,000+ organizations use *Relativity*.⁸

Another type of tool clusters information into logical groups for simplified sorting. *Relativity’s Cluster Visualization Circle pack*,⁹ for example, works within *Relativity* (and other e-discovery tools) to create groups of conceptually similar documents. To create these groups, of course you first need to get your ESI into *Relativity’s* software. You then select documents of interest, which may be documents from several sources or documents within a particular data range, and submit those

documents for clustering by *Relativity’s* Analytics engine. The engine identifies what it deems to be the most logical groupings of documents and places them in clusters. (See *Figure 2*.)

These visual clusters can be used for many purposes, but perhaps the most common use is to help put similar documents into groups to streamline the review process. It will be easier and faster for a single reviewer to look at a grouping of 500 similar documents than to review those same documents interspersed among 4,000 other documents. That single reviewer also is more likely to make consistent decisions about how those documents should be treated than 50 different reviewers would.

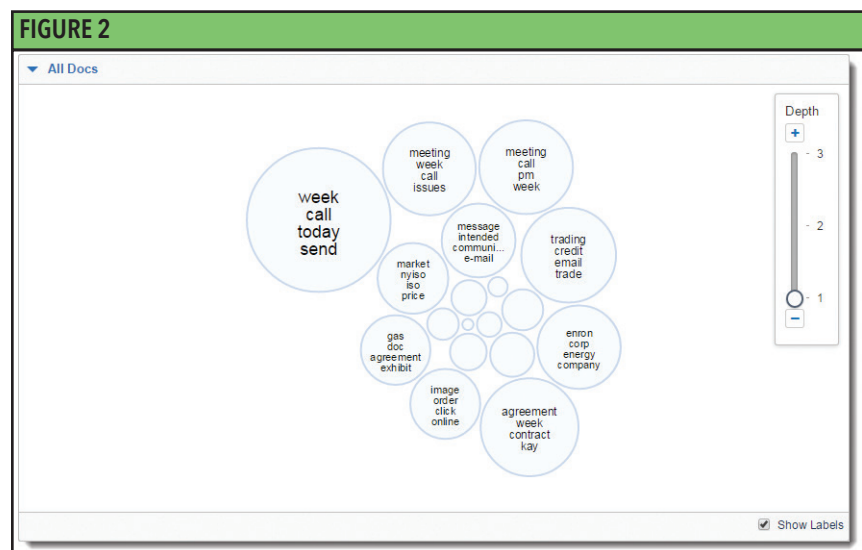
Visual clusters also can identify groups of documents that may not need to be reviewed at all, or groups that counsel determine, after

reviewing a sample, are not pertinent to the issues in the lawsuit.

Clustering groups of conceptually similar documents further allows counsel to identify potentially fruitful areas of inquiry that had not previously been considered. With this approach, counsel can unearth new defenses, claims, potential witnesses, and theories of the case.

Another tool with the same objectives in mind is *Brainspace Discovery’s Concept Cluster Wheel*.¹⁰ The underlying technology is different than that used by *Relativity* – and there is not room here to discuss the merits and shortcomings of each – but the uses to which they can be put are similar. (See *Figure 3*.)

The two tools deliver different results (they do use different underlying technologies, after all), which is precisely the reason to use both tools ▶



THE USE OF DATA VISUALIZATION TOOLS IS NOT YET COMMONPLACE AMONG LAW FIRMS AND LEGAL DEPARTMENTS. . . THE USE IS EXPANDING, HOWEVER, AND RAPIDLY.

rather than just one or the other. For example, a firm might start with ESI housed in Relativity and then port portions of that data to Brainspace to evaluate it using the Concept Cluster Wheel. You could then port the results of your examinations back into Relativity, to make use of them there.

Communication Analysis,¹¹ another Brainspace visualization tool, displays networks of communications, such as email and instant messaging. (See Figure 4.)

With Communication Analysis, communications can be displayed by person, as in the accompanying example, or by domain. You can choose whether to show 'to,' 'cc,' 'bcc,' or any combination of those, and whether to show incoming traffic, outgoing traffic, or both. You can zoom in or out, and you can filter the results

in numerous ways. If you are trying to figure out who communicated with whom, about what, and when, a tool like this can be a great asset.

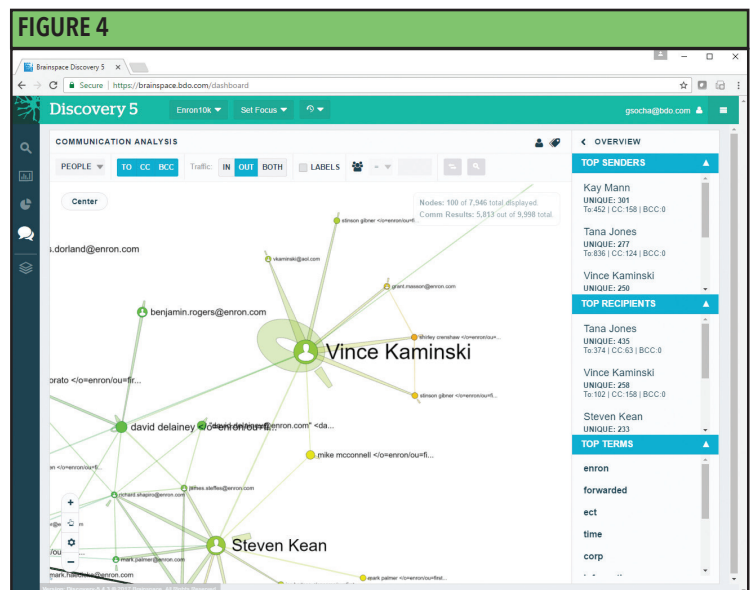
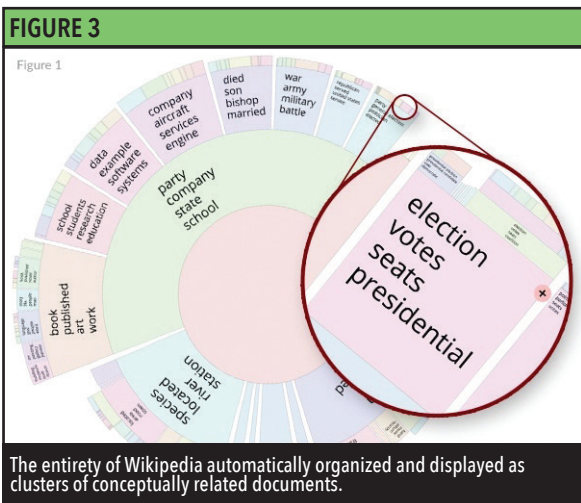
Tableau has been available for over 15 years and is widely used in the larger data analytics world, not just the e-discovery market. Tableau can be used to show clients what types of files they have in their data, sometimes even before the data has been preserved. That information can be displayed with circles, where the larger the circle, the more files of that type were in the population examined, or by file size, file count, and file size by custodian. All are shown in the accompanying example. (See Figure 5.)

Tableau helps you better understand how much data has been filtered out at different stages of the review process. (See Figure 6.)

Tableau can also help you better understand the effectiveness of different key words and combinations of key words. This enhances your ability to decide which key words to use, and which to dispense with, as you are evaluating ESI, making decisions about search strategies, and negotiating with opposing counsel. (See Figure 7.)

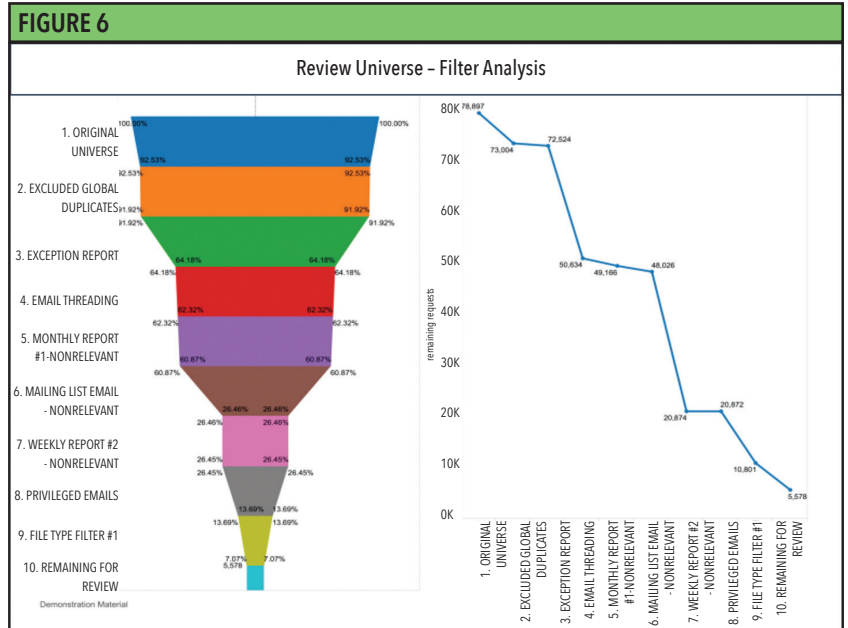
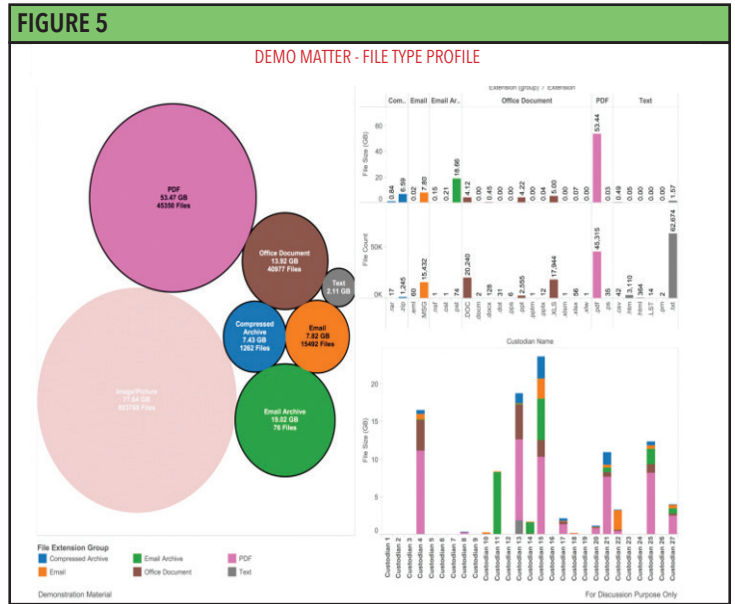
With this article, I have only touched on the myriad ways in which data visualization can be used to enhance the ability of counsel and their clients to work with and understand the ESI they must contend with in the matters they handle. I also highlighted only a few of the many data visualization tools that can be used toward these ends.

The use of data visualization tools is not yet commonplace among law firms and legal depart-



ments (Relativity's numbers notwithstanding). The use is expanding, however, and rapidly. You may not have seen examples like the ones in this article yet, but expect to see them soon – and not just in articles and at conferences but in courtrooms as well.

– **GEORGE SOCHA** IS THE COFOUNDER OF EDM (THE LEADING E-DISCOVERY STANDARDS ORGANIZATION, NOW HOUSED AT DUKE LAW SCHOOL) AND DIRECTOR OF BDO CONSULTING'S FORENSIC TECHNOLOGY SERVICES. THIS IS THE FIRST IN A NEW E-DISCOVERY SERIES THAT EDM WILL PRODUCE FOR *JUDICATURE*.



	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8	Term 9	Term 10	Term 11	Term 12	Term 13	Term 14	Term 15	Term 16	Term 17	
Term 1	1,187																	163
Term 2	1,187	1																154
Term 3		1	1															0
Term 4			260	1														22
Term 5				76	116													7
Term 6				116	65	65	58	78	28	36								19
Term 7				370	388	122	58	78	106	29	21							125
Term 8				592	437	132	76	106	263	28	24							109
Term 9				48	33	35	25	29	36									6
Term 10				84	35	40	36	21	21	34	22							19
Term 11				0	0													0
Term 12				36	33	0	30	11	23	33	22	11	17					22
Term 13				53	115	0	5	2	4	51	50	2	9					53
Term 14				2	1	0	1	2	1	1	1	1	0					1
Term 15				1														1
Term 16				163	184	0	27	7	19	125	109	6	19					1
Term 17				11	3	0	2	2	2	2	1	1	1					1
Term 18				62	73		9	2	36	34	4	6						57
Term 19				98	54		57	12	46	50	52	31	4					20
Term 20				5	3	1	2	0	4	2	1	1	1					1
Term 21				1														1
Term 22				214	116		42	12	36	222	139	13	7					110
Term 23				2	1		1	1	1	1	1	1	1					1
Term 24				1	0													1
Term 25				0	1	0												0
Term 26				25	17		19	11	15	11	16	12	6					0
Term 27				0	0													0
Term 28				0	0													0
Term 29				0	0													0
Term 30				0	0													0
Term 31				0	0													0
Term 32				0	0													0
Term 33				0	0													0

- 1 Data analytics (DA) definition, TECHTARGET (December 2016), <http://searchdatamanagement.techtarget.com/definition/data-analytics>.
- 2 Data visualization definition, TECHTARGET (May 2012), <http://searchbusinessanalytics.techtarget.com/definition/data-visualization>.
- 3 EDWARD TUFT, THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION (1983).
- 4 Skip Walter, Attenex Patterns History – The Critical First Year, ON THE WAY TO SOMEWHERE ELSE BLOG (Jan. 26, 2012), <https://skipwalter.net/2012/01/26/attenex-patterns-history-the-critical-first-year/>.
- 5 See RELATIVITY, <https://www.kcure.com/relativity/>.
- 6 See BRAINSPACE DISCOVERY, <https://www.brainspace.com/products/discovery/>.
- 7 See TABLEAU, <https://www.tableau.com/>.
- 8 Relativity Overview, RELATIVITY, <https://www.kcure.com/relativity/ediscovery-products/overview/>.
- 9 Cluster Visualization, RELATIVITY, https://help.kcure.com/9.2/Content/Relativity/Analytics/Cluster_visualization.htm
- 10 Brainspace Discovery's Concept Cluster Wheel, BRAINSPACE DISCOVERY, <https://www.brainspace.com/blog/e-discovery-searching-for-the-narrative/>.
- 11 Brainspace Discovery, <https://www.brainspace.com/products/discovery/>.