Copyrighting Big Data

Michael Mattioli

Abstract: A new kind of science is transforming diverse industries—from medicine, to energy, to online services. This revolution is fueled by “big data”—vast and varied collections of information that hold new insights about nearly every avenue of human endeavor. This Article examines one of many intellectual property law questions prompted by the swift spread of these activities: To what extent, if any, can collections of big data be protected under copyright law? The label “big data” has come to imply indiscriminate and highly automated data collection practices that yield impartial representations of facts. If these characterizations are true, then copyright protection for big data should be thin, and perhaps should only get thinner as datasets expand. Through a series of original case studies, this Article demonstrates that the opposite is true: Ironically, the very methods and practices that make big data so useful can also infuse both the data collection process and even the underlying data with subjective human judgments. Stated differently, making use of big data often requires “selecting, coordinating, and arranging” the data in creative ways. Surprisingly, big data compilations may be more likely to satisfy the prerequisites for copyrightability than canonical factual compilations. This Article lays the empirical foundation for this conclusion, and explores its economic and political implications.

‡‡‡‡‡‡‡‡‡‡‡‡ Associate Professor of Law, Indiana University Maurer School of Law.