

## Data Pools

### Précis

Michael Mattioli  
Associate Professor  
Indiana University Maurer School of Law  
[michaelmattioli.org](http://michaelmattioli.org)

This project examines the intellectual property policy implications of “big data pools”—a new type of collective rights organization (CRO) swiftly taking form in various industries.

This project is motivated by a longstanding policy problem: how to encourage the collection and distribution of data while preserving widespread access to it. Scholars first recognized this problem decades ago as a straightforward “free-rider” challenge: economic incentives to collect and distribute data are arguably dampened because data is costly to gather and easy to copy.<sup>1</sup> In light of the thin protection that data generally receives under traditional IP law, Congress has considered but repeatedly rejected at least ten proposals to solve this problem through *sui generis* data protection—i.e., intellectual property-like protections designed specifically for data.<sup>2</sup> These proposals failed, in large part, due to widespread concerns that imbuing data with intellectual property-like ownership rules could severely chill scientific inquiry, and might slow the overall pace of scientific discovery.

The recent formation of big data pools may herald a new and fundamental shift in this debate. Like earlier collective rights organizations (e.g., patent pools), these collectives appear to privately craft intellectual property-like protections through contracts. Specifically, they appear to induce their members to contribute valuable subject matter (data) by offering them a number of exclusive rights and shared privileges in return. If the foregoing characterization is accurate, then perhaps the proper subject of debate should shift away from *sui generis* data protection and toward the government’s power to foster the development of private data-sharing collectives.

This project explores this intriguing possibility through a set of ethnographic case studies that examine the structure and governance of big data pools and some of the challenges they face. The groups studied include: CancerLinQ, The Data

---

<sup>1</sup> A form of the same basic problem was at issue in *INS v. AP* 248 U.S. 215 (1918).

<sup>2</sup> Database Investment and Intellectual Property Antipiracy Act of 1996, H.R. 3531, 104th Cong. (1996); Collections of Information Antipiracy Act, H.R. 2652, 105th Cong. (1998); Consumer and Investor Access to Information Act, H.R. 1858, 106th Cong. (1999); Collections of Information Antipiracy Act, H.R. 354, 106th Cong. (1999); Database and Collections of Information Misappropriation Act, H.R. 3261, 108th Cong. (2003); Consumer Access to Information Act of 2004, H.R. 3872, 108th Cong. (2004); Database Investment and Intellectual Property Antipiracy Act, H.R. 3531, 104th Cong. (2004); Collections of Information Antipiracy Act, H.R. 2652, 105th Cong. (2005); Collections of Information Antipiracy Act, H.R. 354, 106th Cong. (2006); Consumer and Investor Access to Information Act, H.R. 1858, 106th Cong. (2006).

Alliance Collaborative, The Biomarkers Consortium, Project Datasphere, and the Open Ag Data Alliance. From this study, this project identifies a range of new policy measures that could foster private ordering solutions to the problem of encouraging data collection while preserving access. These include possible changes to antitrust, privacy, and consumer protection laws, as well as government-led efforts to push for standards and to establish and administer data pools.

In addition to making original empirical contributions, this project seeks to show how the ends of intellectual property policy can sometimes be meaningfully furthered indirectly through policies that fall outside the traditional IP framework.

Michael Mattioli  
August 5, 2014