Press Freedoms in the United States 2021

Data journalism wins some shelter, but still faces threats

By Legal Fellow Grayson Clary

May 24, 2022

The **experience of Josh Renaud**, a reporter for the St. Louis Post-Dispatch, doesn't look much like most of the incidents cataloged by the U.S. Press Freedom Tracker in 2021. Renaud drew the ire of the state of Missouri not for documenting law enforcement at a protest or obtaining confidential government documents but for something most of us do everyday — visiting a website. But as unusual as his story was, it highlights the importance to press freedom of shoring up protections for data journalists.

As the Reporters Committee has often highlighted, contemporary journalism runs on digital investigative techniques. One reporter might **scrape a site** to aggregate data and uncover trends that aren't visible to a naked-eye user; another might use **search operators** to surface files that a website owner accidentally exposed to the public. In Renaud's case, he inspected the published source code of a website run by Missouri's Department of Elementary and Secondary Education, learning in the process that the state had unintentionally published the Social Security numbers of thousands of teachers. After quietly flagging the flaw and giving state officials a chance to repair it, the Post-Dispatch **published its story** on the agency's mistake.

But exactly because these techniques can expose newsworthy shortcomings that some might prefer to keep secret, they can draw legal threats. Missouri Gov. Mike Parson attempted (without success) to have Renaud brought up on

criminal hacking charges for his reporting; as we've written before, that tactic **isn't as rare** as you might think. In the view of too many powerful officials and firms, you need their permission to notice what they've chosen to put online — or else you'll be labeled a hacker, no better than someone who cracks a bank account password.

> **"In the view of too many powerful officials and firms, you need their permission to notice what they've chosen to put online — or else you'll be labeled a hacker."**

Many — the Reporters Committee **included** — hoped the U.S. Supreme Court would help put that theory to rest in *Van Buren v. United States*, the Court's **first significant encounter with the federal anti-hacking law**, the Computer Fraud and Abuse Act. In the lead-up to the case, its stakes were often framed in terms of a contest between what computer crime scholar Professor Orin Kerr **has called** "code-based" and "contract-based" interpretations of the statute. Do you violate federal law by violating the private expectations that a website owner sets out, as when they adopt terms of service that purport to prohibit collecting data on their site? Or does the law only punish those who engage in what looks classically like hacking — bypassing some technical barrier to accessing a computer, as by stealing or circumventing a password?

As we've often argued, the first reading is an exceptionally dangerous one that would let private parties decide whether to criminalize important newsgathering online. After all, website owners routinely purport to prohibit journalists and researchers from noticing information that anyone with a web browser can view — as if a shop owner could use the threat of prosecution to police who is and isn't allowed to look at a window display from a public street. And to a substantial degree the **Supreme Court agreed** with our view, warning that reading the law to prohibit lawful access to a computer for an unauthorized purpose would "criminalize everything from embellishing an online-dating profile" to routine "journalism activity." But its road to that result was somewhat cryptic, purporting to leave unresolved whether determining if someone is entitled to access a computer in the first place turns only on "technological" realities or limitations "contained in contracts or policies" as well. That left open important questions about the decision's impact for data journalists, gaps the lower courts are only just beginning to fill in.

Some early signs are promising. In a closely watched decision in *hiQ Labs Corp. v. LinkedIn*, the U.S. Court of Appeals for the Ninth Circuit **concluded** that *Van Buren* reinforced its view that scraping a publicly available website doesn't violate the CFAA — even if the site owner would prefer you stop. (The Reporters Committee had filed a **friend-of-the-court brief** on behalf of a coalition of news organizations urging that result.) But the decision is the first word, not the last, on the issue. And as Renaud's case highlights, federal law is only a piece of the puzzle. What about state criminal statutes that may or may not line up neatly with the CFAA's terms? Or state torts like trespass that website operators have attempted to use to enforce digital property lines? What about techniques beyond scraping that may lead a journalist to resources a site owner didn't realize were public at all, as Renaud's use of his browser's "view source" functionality did?

As these issues work their way through the courts, we'll continue advocating for the First Amendment rights of journalists to gather data where they find it, online or off. But as this Tracker story highlights, work remains to be done — and this is a press freedom trend to watch.

REPORTERS COMMITTEE
FOR FREEDOM OF THE PRESS

Grayson Clary is the Stanton Foundation National Security/Free Press Fellow at the Reporters Committee for Freedom of the Press and a member of the advisory board for the U.S. Press Freedom Tracker.

We're sending this from our homes, but you can usually find us at:
1156 15th St. NW, Ste. 1020, Washington, D.C., 20005