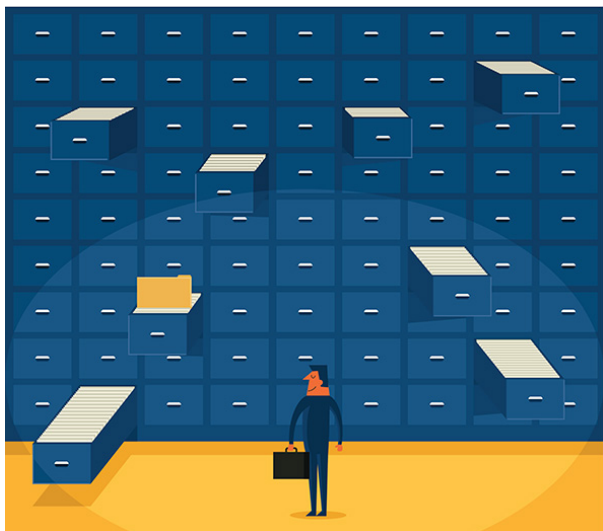


Philanthropy & Funding

# Unlocking the Potential of Open 990 Data

As the movement to expand public use of nonprofit data collected by the Internal Revenue Service advances, it's a good time to review how far the social sector has come and how much work remains to reach the full potential of this treasure trove.

By [Cinthia Schuman Ottinger](#) & [Jeff Williams](#) | Sep. 8, 2022



(Illustration by iStock/sorbetto)

The social sector is using big data to enhance nonprofit transparency and knowledge more than ever before, and the opening of the Form 990 has made an essential contribution. For years, the vast universe of data collected by the US Internal Revenue Service (IRS) via Form 990—including information on nonprofit finances, the size and scope of the US nonprofit sector, the level of charitable giving, employment and volunteering, types of nonprofit activities, and nonprofit organizational and governance structures—was largely inaccessible to the public due to enormous costs and inefficiencies. But the release of [landmark research](#) in 2013, combined with a

concerted campaign by the Aspen Institute and partners such as Candid and the nonprofit research centers at the Urban Institute, Indiana University and Johns Hopkins University, made a strong case for greater transparency and efficiency. This led to a series of legal and policy victories, and ultimately to federal legislation in 2019 mandating electronic filing of all 990s and the release of this information in a publicly available, machine-readable format. The IRS is currently completing its rollout of the new law.

The scale and speed of knowledge gained from Form 990 data would have been unattainable prior to its fully free and public release, according to [a recent study](#) by the Dorothy A. Johnson Center for

Philanthropy, which points to dozens of breakthroughs where open Form 990 data enabled new insights and practices.

Yet despite these breakthroughs, the social sector has only begun to scratch the surface of open 990 data's capabilities. Due to limitations inherent in the data and the way it is currently made available, it's far from reaching its potential. Making more robust use of open 990 data requires that nonprofits, foundations, researchers, and the IRS and federal government alike commit to sustained action.

## A Key to New Insights and Practices

Organizations have employed open Form 990 data in numerous ways, including to:

- **Create new tools for donors.** For instance, the Nonprofit Aid Visualizer, a partnership between Candid and Vanguard Charitable, uses open 990 data to find communities vulnerable to COVID-19, and help address both their immediate needs and long-term recovery. Another tool, COVID-19 Urgent Service Provider Support Tool, developed by the consulting firm BCT Partners, uses 990 data to direct donors to service providers that are close to communities most affected by COVID-19.
- **More efficiently prosecute charitable fraud.** This includes a campaign by the New York Attorney General's Office that recovered \$1.7 million from sham charities and redirected funds to legitimate groups.
- **Generate groundbreaking findings on fundraising, volunteers, equity, and management.** A [researcher at Texas Tech University](#), for example, explored more than a million e-filed 990s to overturn long-held assumptions about the role of cash in fundraising. He found that when nonprofits encourage noncash gifts as opposed to only cash contributions, financial contributions to those organizations increase over time.
- **Shed light on harmful practices that hurt the poor.** A [large-scale investigative analysis](#) of nonprofit hospitals' tax forms revealed that 45 percent of them sent a total of \$2.7 billion in medical bills to patients whose incomes were likely low enough to qualify for free or discounted care. When this practice was publicly exposed, some hospitals reevaluated their practices and erased unpaid bills for qualifying patients. The expense of mining data like this previously made such research next to impossible.
- **Help donors make more informed giving decisions.** In hopes of maximizing contributions to Ukrainian relief efforts, a record number of donors are turning to resources like Charity Navigator, which can now use open Form 990 data to evaluate and rate a large number of

charities based on finances, governance, and other factors. At the same time, donors informed by open 990 data can seek more accountability from the organizations they support. For example, anti-corruption researchers scouring open 990 data and other records uncovered donations by Russian oligarchs aligned with President Putin. This pressured US nonprofits that accepted money from the oligarchs to disavow this funding.

Regular, robust use of open big data in the social sector could employ text analysis and natural language processing to shed light on best practices in fundraising, governance, and other important nonprofit functions. It could mash together disparate datasets to reveal insights that help solve intractable community problems, such as overlapping nonprofit and philanthropic resources with data on community needs. It could also permit charity officials to quickly identify, track, and stop charitable misconduct and fraud, within states and across state lines. And it could provide policy makers and the public with accurate information on government relief funding delivered via grants to the social sector, by state and zip code, to identify gaps in assistance.

## Barriers to Potential

Yet despite its patent potential, barriers to access and use remain. For one, the social sector often lacks the coordinated infrastructure needed to consistently use it. It has yet to fully connect relevant data sets to core constituencies: not just to donors, but also to practitioners, researchers, journalists, charity regulators, policy makers, and funders who could use data-informed knowledge to propel their work.

Another concern is ease of use. The total size of open Form 990 data means it's generally impractical for individuals acting alone to download, manage, or analyze the information. The sector needs data files that are already structured for use in familiar software (such as Excel or Tableau), easily searchable, and readily available on both a trend-level and individual organization basis. While the breadth and depth of open Form 990 is revolutionary, it is a very "high effort" dataset for individual users who want direct access.

A further challenge is that the IRS releases data on an irregular schedule. More than a year can pass between the time an organization files a 990 and the first appearance of the data in the open data repository. And even when the data is ready for release, the IRS doesn't make public sets of data on a predictable or particularly timely schedule. This makes it difficult for foundations and nonprofits seeking to respond to changing needs in the field to get the information they need.

Finally, 990 forms themselves have both incomplete information and mismatched fields between nonprofit filers by size. Even though government now mandates electronic filing, some nonprofits

provide required information in alternative formats that make electronic analysis more difficult. A common example is a foundation that provides its grantee list in an unstructured text attachment, instead of in the form's structured fields, or a nonprofit that doesn't cleanly list its directors and compensation information. Mismatched fields lead to inconsistencies in field naming and field definitions, making quick comparisons between data sets time consuming. In addition, even when the four primary 990 forms (990, 990-EZ, 990-N, and 990-PF) ask the same basic questions about revenue or expenditures for nonprofits regardless of size, the questions and the answers appear in different parts, sections, and line numbers of each form.

These challenges are not insurmountable. In fact, they point toward potential solutions that encourage collaboration. The social sector must work together to build the data infrastructure across institutions and members of the public, as well as collaborate with the IRS to reduce barriers and create a public system that bolsters data infrastructure.

## Suggestions for Action

Nonprofits, foundations, researchers, and the IRS all have roles to play. First, the social sector needs to promote educational resources that inspire and instruct novice practitioners on how to incorporate open data into their solutions to complex social problems. This includes highlighting, celebrating, and sharing projects that successfully tap into open Form 990 data feed and demonstrate possible uses. Examples include the [Dorothy A. Johnson Center for Philanthropy study](#) and [companion repository of open Form 990 use cases and public reports](#); tools for accessing the 990 data, such as the data and analysis programming scripts [published by researchers for free through public access on GitHub](#); and the [Aspen Institute's brochure](#) with tips for organizations now filing electronically.

The sector also needs to move beyond the existing, informal network of volunteers, researchers, and commercial and nonprofit data providers who create and publish individual tools. These intermediaries provide users (that know where to look!) with access to parts of the 990 universe, but there isn't a comprehensive, coordinated tier of supports with clear access points and a regular publication schedule to help reach the broadest possible audience. No group regularly publishes and maintains common and comprehensive "grab and go" data sets; nor is there a widely available raw data set with cumulative or summary statistics providing even basic trend-line information about the sector, large nonprofit subsectors, or nonprofit organizations by region.

In short, the sector needs to cooperate and coordinate to create a freely available and comprehensive data ecosystem. This would provide nonprofit and commercial data users multiple places for

engagement, while increasing the accessibility and utility of the data files for custom “mashups” with other related datasets.

Even with augmentation, a sector that accounts for over 10 percent of the nation’s private employment and provides more jobs than the manufacturing industry shouldn’t have to rely solely on a network of volunteers for access to basic summary data. Across the federal government, departments and agencies need to step up and recognize the importance of nonprofits to the financial output of the United States. For example, the IRS can take steps to improve the utility, consistency, and accessibility of existing open Form 990 data by providing a regular schedule of data releases, posting a data dictionary to help users identify fields more quickly within each data set, and asking the research community for suggestions about improving machine-readable data accessibility and categorization in text fields.

In addition, Congress and the Treasury Department need to provide both financial and institutional support for the IRS tax-exempt staff—especially the data-processing team—by increasing the staff size and funding for the charitable sector regulators. Because the tax-exempt sector generally doesn’t generate revenue for the US government, there are fewer resources available to it, even though accessible nonprofit data would help boost compliance efforts.

Finally, the single best action to advance knowledge throughout the nonprofit sector would be to require every federal agency to disaggregate its data for the nonprofit sector whenever it publishes data at a sectoral or industry level. This includes requiring the US Department of Labor to release nonprofit employment and wage data quarterly on a national, state, and county/metropolitan area basis, broken down by fields. This step, as described in a [letter to the US Secretary of Education](#) signed by hundreds of nonprofit organizations and scholars, is essential for nonprofit planning and policymaking, and would make nonprofits and foundations as visible as the manufacturing or retail sectors in federal statistical publications.

The future health, resilience, and effectiveness of the social sector hinges on the ability to unlock the potential of big data. Building and maintaining the information infrastructure we need as quickly and efficiently as possible is imperative to creating new insights into the health of the nonprofit sector, and to deepening the pathways along which massive quantities of funds, resources, and people engage with organizations and communities. By matching the Form 990 data to other existing public datasets, and clarifying the underlying data structure and tools for data access and validation, open Form 990 data can live up to its potential and support a thriving civil society.

*Read more stories by [Cynthia Schuman Ottinger](#) & [Jeff Williams](#).*

---



Cinthia Schuman Ottinger is deputy director for philanthropy programs at the [Aspen Institute's Program on Philanthropy and Social Innovation](#). She leads the Nonprofit Data Project, which spearheaded the passage of the “open” 990 data law, federal legislation to shed greater light on the vast US nonprofit sector, and convenes the Nonprofit Open Data Collective.



Jeff Williams is the director of the community data and research lab at Grand Valley State University's [Dorothy A. Johnson Center for Philanthropy](#). In this role, he leads a team of researchers, evaluators, database experts, and web developers who provide insights into nonprofit and philanthropic organizations and activities.

If you like this article enough to print it, be sure to subscribe to SSIR!

Copyright © 2022 Stanford University.  
Designed by Arsenal, developed by Hop Studios