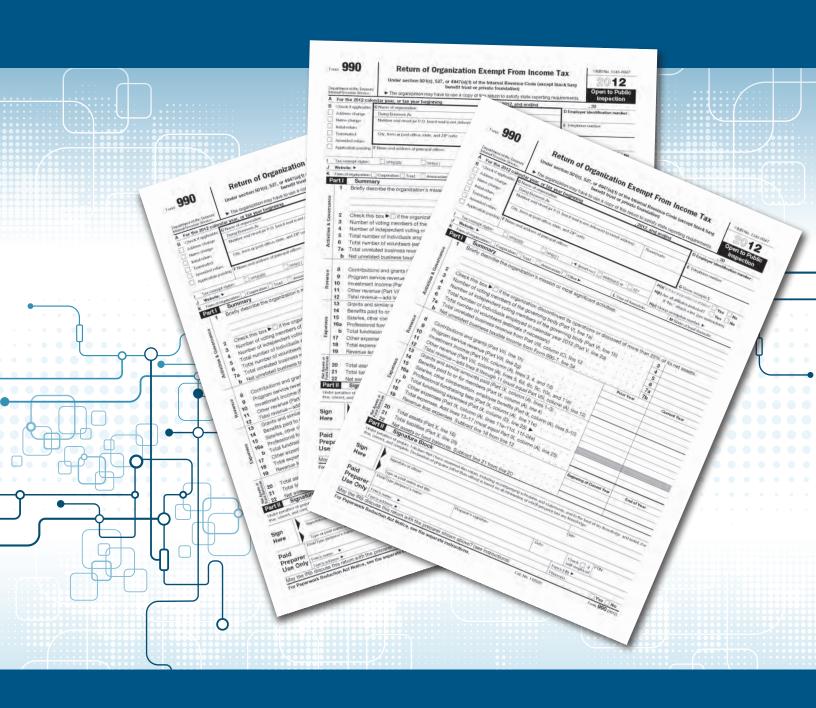THE ASPEN INSTITUTE
Philanthropy & Social Innovation

Information for Impact:
Liberating Nonprofit Sector Data

By Beth Simone Noveck and Daniel L. Goroff

Information for Impact:
Liberating Nonprofit Sector Data

By Beth Simone Noveck and Daniel L. Goroff*

THE ASPEN INSTITUTE
Philanthropy & Social Innovation

About the Program on Philanthropy and Social Innovation

Through convenings, leadership development initiatives, communications, and strategic partnerships, the Aspen Institute's Program on Philanthropy and Social Innovation (PSI) seeks to maximize the impact of social-sector leaders in contributing to the good society at home and abroad.

PSI's Nonprofit Data Project works to promote accessible, accurate, and current information on the U.S. nonprofit sector in partnership with the nation's top nonprofit data providers — the Foundation Center, GuideStar, Center for Civil Society Studies at Johns Hopkins University, Center on Philanthropy at Indiana University, and the Center on Nonprofits and Philanthropy at the Urban Institute. The Nonprofit Data Project wishes to thank its supporters, the Bill & Melinda Gates Foundation, the Charles Stewart Mott Foundation, and the Goldhirsh Foundation.

The Program on Philanthropy and Social Innovation also hosts the Aspen Philanthropy Group, an agenda-setting body of foundation, public, and private-sector leaders at the cutting edge of change, and it spurs partnerships and collaborative action among them. Leadership-development initiatives include the American Express Foundation-Aspen Institute Fellowship for Emerging Nonprofit Leaders, the Aspen Philanthropy Seminar, and the Seminar for Mid-America Foundation CEOs. PSI's policy work includes the Impact Economy Initiative, which seeks to strengthen and expand the field of impact investing. The Global Philanthropy Forum serves as a key partner in many of PSI's efforts.

The Aspen Institute's mission is twofold: to foster values-based leadership, encouraging individuals to reflect on the ideals and ideas that define a good society, and to provide a neutral and balanced venue for discussing and acting on critical issues. The Aspen Institute does this primarily in four ways: seminars, young-leader fellowships around the globe, policy programs, and public conferences and events. The Institute is based in Washington, D.C.; Aspen, Colorado; and on the Wye River on Maryland's Eastern Shore, and has an international network of partners.

INFORMATION FOR IMPACT: LIBERATING NONPROFIT SECTOR DATA

TABLE OF CONTENTS

Executive Summary

This report addresses the challenges to obtaining better, more usable data about the nonprofit sector to match the sector's growing importance. In 2010, there were 1.5 million tax-exempt organizations in the United States with $1.51 trillion in revenues. Through the Form 990 in its several varieties, the Internal Revenue Service (IRS) gathers and publishes a large amount of information about tax-exempt organizations. Over time, versions of the Form 990 have evolved that collect information on governance, investments, and other factors not directly related to an organization's tax calculations or qualifications for tax exemption. Copies of these returns are available one at a time from the filers or from other sources. The IRS creates image files of Form 990 returns and sells compilations of them to the subscribing public for a fee. Several institutions, particularly GuideStar, the Foundation Center, and the National Center for Charitable Statistics (NCCS) at the Urban Institute, use this IRS data to analyze and present information about individual nonprofits and about the sector as a whole.

Like other important data collected by governments, information contained in the 990s could potentially be far more useful if it were not only public but "open" data. Open data are data that are available to all, free of charge, in a standard format, published without proprietary conditions, and available online as a bulk download rather than only through single-entry lookup. Making the Form 990 data truly open in this sense would not only make it easier to use for the organizations that already process it, but would also make it useful to researchers, advocates, entrepreneurs, technologists, and nonprofits that do not have the resources to use the data in its current form. We argue that open 990 data may increase transparency for nonprofit organizations, making it easier for state and federal authorities to detect fraud, spur innovation in the nonprofit sector and, above all, help us to understand the potential value of the 990 data.

However, as also is the case for the 990s, "liberating" government data from paper-based technologies and processes is neither straightforward nor free. While myriad benefits may justify the costs, as we argue in this case, there is a need for practical strategies to overcome technical, political, and cultural impediments. The paper explores the costs and benefits of four avenues, which are not mutually exclusive, for achieving open 990 data:

1. A legislative mandate for e-filing: Congressional action to mandate electronic filing combined with guidance on how electronic data should be released.

2. IRS initiative: A major project by the IRS to turn Form 990 data into open data.

3. Third-party platform: A third party that would take on the task of creating an open database from Form 990 data.

4. A priori electronic filing: A variation on (3) in which nonprofits would, when filing with the IRS, simultaneously send their electronic 990 information to a third party responsible for database creation and maintenance.

After reviewing the pros and cons of these options, we recommend a two-track strategy: Pursuing the longer-term goal of legislation that would mandate electronic filing to create open 990 data, and pursuing a shorter-term strategy of developing a third-party platform that can demonstrate benefits more immediately. The paper concludes with specific recommendations for rapid development of an initial database combined with contests, challenges, and other approaches to build a community willing to examine the quality of 990 data and build innovative tools that draw upon 990 and other available data to create a better understanding of the nonprofit sector than we currently have and that we so desperately need.

THE 990 OPPORTUNITY

HIGHLIGHTS

- To be fully usable, data need to be "open." Data need to be accessible in forms that third parties can use to do analysis and develop software applications.

- Opening data contained in the nonprofit tax returns (Form 990 data) should foster innovation in the nonprofit sector.

- This paper addresses strategies for making Form 990 data open, considering both government and third-party solutions, with the goal of spurring transparency and innovation.

America's nonprofit sector is the envy of the world. Over the past fifty years, this robust sector has grown one hundred fold while giving rise to everything from the civil rights, environmental, and Tea Party movements to revolutions in agriculture, medicine, and the media. Universities, churches, foundations, charities, and social service organizations of all kinds are major drivers of culture, art, religion, science, and altruistic services in the United States.

The role of nonprofits may now become even greater – and subject to greater scrutiny. With a national movement in the U.S. to shrink the role of government, non-governmental organizations (NGOs) will be expected to step in to fill essential needs, not just fill in the gaps left by government programs. That expectation will put pressure on NGOs to develop more comprehensive, effective, and efficient programs, and to release more information about their operations, investments, and impact. Entrepreneurial donors of wealth and influence — not to mention regulators, economists, journalists, and all of us concerned with the state of our country — will demand greater accountability from nonprofits than ever before.

In principle, the best way to provide more information about the nonprofit sector could be to work with the data federal and state governments already collect from nonprofits. Different states have different regulations regarding nonprofit filing. Some are now working together on systems for electronic registration of nonprofits, and some require federal forms to be filed with the state as well. But the most thorough and useful data sources are the different versions of the Form 990 filed with the Internal Revenue Service.

The Form 990 dates back to the Revenue Act of 1943 and has always requested information about a tax-exempt organization's activities. Over time, however, nonprofits have been expected to file more extensive information that gives greater insight into their operations. Most are required to file information about unrelated business income, officers' compensation, and information about governance, finances, and investments that has become increasingly detailed over the years. Academics, independent organizations, and nonprofits themselves have used these data to analyze individual organizations and the nonprofit sector as a whole.

The U.S. Internal Revenue Service collects copious data about the nonprofit sector through the different versions of the Form 990. Every year, approximately 1.5 million registered tax-exempt organizations file one of the Forms 990, [2] but the corpus of Forms 990 data have not been made available in a comprehensive and open format. This trove of data about a sector that pays $670 billion in wages and benefits annually [3] includes financial, governance, and organizational information. [4]

The IRS already devotes significant resources to making 990 data public, as is required by law. But returns are released only as individual image files. This format is useful only for reading about one nonprofit organization at a time. The sector deserves comprehensive and computable data that can be openly aggregated, searched, checked, and analyzed. In short, 990 data needs to be liberated.

Releasing 990 information as "open data" — in a "computable" format that can be electronically analyzed and used for many purposes — would make the entire nonprofit sector more transparent. When government agencies release the data they collect as open data, great things can happen. For example, the U.S. Department of Health and Human Services recently published its database of hospital infection rates online in a computable format.

With this data publicly available for the first time, Microsoft and Google were able to mash it up with mapping data to create an application that shows infection rates for local hospitals across the country. This tool readily allows anyone — from the investigative journalist to the parent of a sick child — to see which hospitals are safest.

Open data can be transformative when third parties use it to create new analyses, visualizations, and mash ups that transform raw data into knowledge. Solving complex challenges requires many people with diverse skills and talents working together. [5] When experts of all kinds have access to open data, it becomes a catalyst for creative problem solving and community innovation. "The real value [of open data] comes from interpretation, analysis, linking-up and reflection — in short, from being used," says one database expert. "Keeping hold of data that you don't use much is wasting potential. By making data open, you enable others to bring fresh perspectives, insights, and additional resources to your data, and that's when it can become really valuable." [6]

Opening up 990 data could provide great opportunities to analyze and improve the nonprofit sector. Researchers are already analyzing the nonprofit sector from economic, geographic, and operational perspectives. [7] A comprehensive source of high-quality data on nonprofits, structured to allow comparisons and analyses across different organizations in the sector, would greatly enhance and accelerate these research efforts. Making IRS data open won't solve every problem; in particular, the fact that many nonprofits file extensions will still pose a challenge to data currency and mistakes entered into forms will still pose a challenge to data quality. However, a comprehensive open data resource would make it possible to:

- Do more extensive, in-depth empirical research on the sector as a whole, including sector-wide issues such as the impact of the economic downturn on nonprofits, the geographic distribution of nonprofit services, and the efficiency of the nonprofit sector in delivering services;

- Combine the 990 data with other datasets, such as those on government spending, to better understand the relationship between public and private dollars in providing social services;

- Query the data to address issues relating to specific nonprofits, such as gaining greater insight into 501(c)(4) organizations that engage in lobbying or finding trends and outliers in executive compensation;

- Recognize fraud early, anticipate abuses, and target enforcement more efficiently and effectively;

- Enable more people and organizations to analyze, visualize, and mash up the data, creating a large public community that is interested in the nonprofit sector and can collaborate to find ways to improve it; and

- Spot issues of both data currency and data quality, and evaluate their impact on our understanding of the sector.

In sum, the data that the IRS collect about nonprofit organizations present a great opportunity to learn about the sector and make it more effective.

These goals are not new. Groups such as the Foundation Center, GuideStar, Urban Institute, Johns Hopkins' Center for Civil Society Studies, and Indiana University's Center on Philanthropy have done data-driven research on the nonprofit sector for years. But there has been no comprehensive set of open data about the sector available to them or the many others who would take advantage of it. Some data resources now available are not computable — that is, they cannot be downloaded and analyzed by other users. Some are not available free of charge. And those data sets that are computable are limited in the kinds of 990 data they include, the number of filers they cover, or both. [8]

Even though the IRS makes Form 990 returns public, it provides no free, online repository where the corpus of data can be downloaded in bulk for re-use (it does provide a download sample of the Forms 990, 990-EZ, and 990-PF.). The obstacles to working with the corpus of Forms 990 data as a whole has made it difficult to compare nonprofit organizations using standard metrics; to analyze issues across the sector, such as the relationship between large donors and board governance; or to mash up IRS data with other data sets, such as data on political contributions.

USA.GOV MEETS USA TODAY
WHEN GOVERNMENT AND THE PUBLIC WORK TOGETHER WITH OPEN DATA: THE FEDERAL REGISTER 2.0 CASE STUDY

In late 2009, the Federal Register, the must-read publication for government policymakers, made ten years of editions available as a bulk data file for free download. In doing so, the National Archives and Records Administration, which publishes the Federal Register, decided to forgo the $17,000 annual subscription price they had previously charged in favor of openness.

Sitting in a coffee shop in San Francisco in April 2010, Dave Augustine, Bob Burbach, and Andrew Carpenter happened upon the "Apps for America" contest, which called for making a useful tool with a government dataset. While they had never heard of it, the Federal Register was the largest file available on the U.S. government's new data portal, Data.Gov. The trio thought they could make the Register look better and be more accessible. Their prototype only won second place in the contest.

But the Federal Register liked it so much they called Dave, Bob, and Andy and commissioned the redesign. These three part-time "citizen coders" (they didn't quit their day jobs) ended up making the actions of the executive branch of government truly accessible to the American people for the first time. Instead of dense, hard to read PDFs, the new Federal Register launched in July 2010 is graphically appealing and searchable. Think USA.gov meets USA Today.

Figure 1 – Federal Register

On another level, the inaccessibility of Form 990 data has made it difficult to know how valuable the data really are. Some experts believe that the usefulness of the 990 data is seriously limited by problems with accuracy, timeliness, and consistency. As Brad Smith, President of the Foundation Center, comments: "Philanthropy is a significant industry, but what other $600 billion industry contends with trend data that is up to three years old?" [9] But in a Catch-22, the only way to determine and improve the usefulness of 990 data is to put it into a more workable form.

The obstacles are not only technical: Even addressing a problem as straightforward as data currency raises tax administration equity problems for a federal agency, which for various reasons, may wish to have a single set of rules for both paper and e-filed returns. The IRS traditionally received 990 data in paper forms that were typed or handwritten. In 2004, the IRS introduced electronic filing, which allows for basic accuracy tests before a return is accepted. Electronic filing is also more efficient: Whereas paper forms have to be scanned into a digital format and only then converted (with difficulty) into something a computer can read, e-filed forms are "born digital" and could be made available without going through such extra steps. But the IRS does not release e-filed data more rapidly than paper data.

Whether a nonprofit files electronically or on paper, there are many challenges to producing clean, reliable, and releasable data. Forms sometimes accidentally include Social Security numbers that should be redacted, for example, though the IRS is legally prohibited from doing so. [10] Corrections made to a form and subsequently filed are not linked back to the original form, since different units of the IRS have different objectives and rules for how they process findings and what they will release publicly. Many nonprofits file extensions or do not file at all, as evidenced by the initial round of exempt-status revocations (one of penalties for chronic failure to file) in 2011. [11] Many also use free-form attachments that lack consistency and accuracy.

Despite the obstacles, the anticipated benefits of achieving current, comprehensive, and usable data about nonprofits could outweigh the difficulties and costs. Moving from a paper-based 990 world of inaccessible and limited data disclosure to free and fully open, computable data will have financial costs, take considerable time, and pose political, cultural, and technical challenges. It would be ideal if the IRS, and every government agency for that matter, were to release its publicly available data in easily used, "machine-readable" formats that can be readily analyzed using computers. The U.S. Government's Digital Government Strategy, released in the spring of 2012, has set machine-readability as a government–wide goal. But the reality is that it can take years for agencies to transition from their current ways of collecting and releasing data to more modern and useful methods. The transition requires new IT systems, new government forms, and new data standards, all of which must be funded by the federal budget.

In some similar situations, outside third parties have worked with government agencies to take on the job of making data accessible and computable. Take the example of state education data. Under the No Child Left Behind Act, each state must release measures of educational quality for its public schools — but the states release the data in very different ways. Third parties have now begun work to make different states' data comparable and give parents and state officials the tools they need to compare public education across the country. A private nonprofit called the Digital Quality Campaign has put together a consortium dedicated to creating more uniform metrics and reports, and a widely used website, GreatSchools.org, has found ways to make different states' data comparable so that parents can evaluate the quality of their local schools on a sound basis. [12]

This report examines the pros and cons of both a centralized government approach and a third-party approach to improving Form 990 data. We also propose that initial work should assess how "opening up" 990 data will increase its value. This pilot work should create a platform with freely available and open 990 data, while also creating incentives for a wider community of users to produce innovations using such data. This demonstration project can help either justify innovation in IRS processes and information technology infrastructure or generate third-party efforts to work with existing IRS data to improve the data, or both.

A Note on Assumptions

When we began work on this paper, we held several assumptions about the best ways to unlock the value of 990 data, which turned out to be unsupported by the evidence we gathered. As we dug in, we realized that in this case like so many other data liberation projects, there are unexpected impediments to openness that must be surmounted. Our starting assumptions included the following:

1. **Congress is the answer.** It's tempting to assume that a Congressional mandate is the only way to open up data about nonprofits. The right legislation would certainly help, but it's not the sole solution. Other possibilities should be explored in tandem with legislative efforts.

2. **All non-private government data should be open now.** "Should" is the operative word here. In an ideal world, public government data would be open data. The Federal Digital Strategy has set that as a goal. [13] But in practice, the obstacles to achieving open government data are many, varied, and often significant.

3. **Silly bureaucracies are the primary impediment.** It would be easy to blame bureaucrats for the lack of more usable government data. In fact, the IRS, which is responsible for Form 990 data, has done a great deal to make Form 990 data available at considerable effort and cost. A combination of technical, policy, and legal issues that are not under the agency's control have prevented more rapid progress.

4. **Transparency is enough.** Open, transparent data about the nonprofit sector are essential. But it will take many other efforts — by third parties, by government agencies, and by individual developers and entrepreneurs — to use the data to help improve nonprofits and make the sector more innovative.

5. **A good report will catalyze the desired change.** Of course, we hope that this report will attract attention to the importance of 990 data. But we recognize that it will take collaboration among many different interested parties to make data about nonprofits more available, usable, accurate, and impactful.

The Forms 990: History and Current Versions

Highlights

- The Forms 990 are a family consisting of the Form 990, Form 990-PF, Form 990-EZ, and Form 990-N. Each has certain thresholds that establish which nonprofit types must file which form.

- The world of the Forms 990 has undergone two recent changes: a major redesign of the Form 990, and the requirement for small nonprofit organizations to file the electronic Form 990-N.

- Any momentum for change within the agency must be consistent with IRS's mission and budget constraints.

Information on nonprofits comes from multiple sources, but much of it supplements the information within the Forms 990. The Internal Revenue Service acts as the tax administrator for the federal government and has regulatory power over the nonprofit community. Among the responsibilities of the IRS in overseeing nonprofits: reviewing applications for the dispensation of tax-exempt status and examining annually filed Forms 990. The Forms 990 require that a nonprofit disclose information about the organization and its transactions, including the nonprofit's major service programs, its officer salaries, and its balance sheet. According to the Internal Revenue Manual (the "IRM"), the primary objectives for examination of annually filed Forms 990 is to determine if the organization is operating in a manner consistent with its exempt purpose, [14] if the forms are complete, [15] if the organization filed all forms and information required of it, [16] and if there are any outstanding taxes the organization must pay (non profits sometimes have associated for-profit activities that are taxable, for example). [17] Violations may result in enforcement ranging from excise taxes to a revocation of tax-exempt status. Some have argued that the information provided within the Forms 990 exceeds that which the IRS needs to conduct its tasks, however, and is meant instead for use by the public.

The family of the Forms 990 includes: the Form 990, which may be used by any nonprofit except private foundations; Form 990-PF, for private foundations specifically; Form 990-EZ, for medium-sized organizations; and Form 990-N, the newest form, which small organizations file electronically. The rules for determining what type of organization must file each form normally depend both on how much the organization earns in revenue and on organization type. Each form varies in length and information requested. (For more details on each, see Appendix 2: The Federal Forms)

The 990 series presents a unique set of challenges for the IRS. In contrast to other filings the agency processes, the IRS collects few if any payments from Form 990 filers. The IRS does charge an initial fee when an organization files for tax-exempt status through Form 1023. The organization is required to pay a $400 fee if it has gross receipts of less than $10,000 a year over a four-year period, and an $850 fee if it has gross receipts greater than $10,000 a year over a four year period. [18]

The IRS has not imposed any fees on organizations that submit Forms 990, even though there has been an increase in the number of nonprofits that file annual returns. In a statement before the House Committee on Ways and Means in 2005, David M. Walker, Comptroller General of the U.S. Government Accountability Office, recognized that IRS caseload had increased and recommended "a full re-examination of the tax-exempt sector in light of the challenges facing the nation in the 21st century." [19]

The kinds of disclosures required in the Forms 990 have changed over the years. The types of questions asked of nonprofits and types of data collected have increased significantly since the first iteration of the Form 990. The most recent substantive changes to the Form 990 occurred in 2008, [20] but the Form 990 has a much longer history that reaches back to the Revenue Act of 1943. The Revenue Act created the requirement for nonprofits to file annual returns. [21] Originally, filings contained considerably less information than they do now. Among the disclosures were gross income, receipts, and disbursements. [22] The legislation exempted certain organizations from filing returns at all. This ended with the first revision of the Form 990. [23]

As the nonprofit community grew, Congress was called on to regulate transactions conducted by nonprofits that were arguably outside the mandate of the nonprofit's original mission. These transactions, such as owning

department stores or selling real estate, were determined to be "unrelated business income." [24] Unrelated business income is now reported on Form 990-T. Under the Revenue Act of 1950, Congress legislated that any unrelated business income earned by nonprofits would be taxed as well. Granting an exemption for a stream of business that would compete with for-profit business would create an "unfair competitive advantage." [25] Foundations would not be covered by legislation until a number of years later, after it was determined that a few were engaged in misconduct. [26] The U.S. Treasury Department created a report that outlined a number of legislative strategies, but some in Congress argued that the suggested action was not strong enough.[27] The Tax Reform Act of 1969 provided a statutory definition for private foundations, and made them a distinct nonprofit category. [28]

The Pension Protection Act of 2006 (PPA), while not directly about the nonprofit community, brought about a number of changes to the sector. The PPA required organizations that earned $25,000 or less in income to file an electronic annual notice, thus creating the need for the Form 990-N.[29] In testimony before the House Committee on Ways and Means, George K. Yin, Chief of Staff of the Joint Committee on Taxation, recommended a proposal that would "require small charitable organizations to file short annual returns," because larger returns would be overly burdensome to small nonprofits. [30] The PPA also added a number of new nonprofit types to the Internal Revenue Code (IRC) [31] and created a new enforcement tool that would allow for the revocation of tax-exempt status for any organization that failed to file a return for three consecutive years. [32] Prior to this, nonprofits that failed to file their returns could only be penalized for failure to file a return. [33]

After the enactment of the PPA, Senate Finance Committee member Charles Grassley stated, "Both the House and Senate tax writing committees have expressed interest in the scope of charity and whether §501(c)(3) organizations are providing a sufficient public benefit by conducting charitable programs commensurate in scope with their resources."[34] This sentiment is reflected as Congress continues to debate such matters, including at recent hearings on regulatory issues and reporting structures of public charities conducted by the Oversight Subcommittee of the House Ways and Means Committee. [35]

In 2008, the Form 990 underwent its first major redesign since 1979. [36] This change was partly motivated by "the increasing size, diversity, and complexity of the exempt sector," and changes were designed to reflect the need for "enhancing transparency, promoting tax compliance, and minimizing the burden on the filing organization." [37] The IRS released a draft of the new Form 990 for comment and received over 700 responses. After taking these responses into account, the IRS released a second draft and received an additional 120 comments. [38] This led to the final iteration of the Form 990, which is the one that is currently used. This revised Form 990 adjusted how compensation for officers was reported; added specific governance questions, such as whether there is a policy in place for potential conflicts of interest; made new financial information requests; and made numerous other significant changes to reporting (e.g., foreign activity reporting, noncash contributions, balance sheet detail, hospital reporting, tax-exempt bond reporting, and related organization reporting).

The most recent version of the Form 990 also includes questions on foreign investments for organizations with investments valued at $100,000 or more. This is a change from previous iterations that only requested foreign investment information if the organization had revenues or expenses attributable to foreign investments that were greater than $10,000. Other changes include adjustments to the balance sheet portion and activities involving joint ventures.

What Is Open Gov Data?

Highlights

- Information must be free from legal and technical restrictions to be most useful.

- Opening data means making information downloadable in bulk and in machine-readable formats for anyone to use free of charge.

What some in the blogosphere call "data hugging disorder" [39] — the tendency for governmental officials, companies, and researchers to hoard data as "mine" and refuse to make it publicly available is giving way slowly to an understanding of the value of open data. Progress toward making federal data open has been a goal of the "Open Government Initiative." President Obama signed the 2009 Memorandum on Transparency and Open Government on his very first day in office, declaring that "information maintained by the Federal Government to be a national asset" and calling for disclosure of government "information rapidly in forms that the public can readily find and use." [40]

All of the Open Government Data You're Already Using

Government data are best used when you don't even realize you're using them. Researchers and analysts have long taken advantage of government data, but entrepreneurs have also successfully used these data to add value and create business models that take advantage of these public resources. Below are some examples of government data that are used by entrepreneurs.

Weather Data: the National Ocean and Atmospheric Administration freely and publicly provides weather and forecast data online. This includes local climatological data, which not only power weather apps, but provide the backbone for such services as the Weather Channel.

Global Positioning System Data: the GPS data we use to get from work to home were made available for civilian use by President Ronald Reagan, who saw the impact these data could provide as a public good.

Restaurant Health Code Violations: local data apps combine city health code violation feeds with Foursquare or other mapping services to help citizens find the rating of a restaurant before they sit down to eat.

Public Transit Information: cities have unlocked the data on when public transit runs and to where, making bus and subways easier to catch than ever before.

Your Health Records: through an effort run by the Department of Veterans Affairs and Department of Health and Human Services, citizens can now download their own personal health data. This takes these data out of the hands of our doctors and gives them to its owners, making it easier than ever before to provide data to new or emergency health care providers.

Figure 2 – Examples of Open Government Data

Since then, 172 federal agencies have contributed to Data.Gov, the national data portal of the United States. In addition, 34 states, 15 U.S. cities, and 30 foreign countries [41] have launched "data dot gov" websites to make public information freely accessible, searchable, and downloadable. Through these websites, the public can access government data without intellectual property restrictions [42] and in formats they can use to analyze, visualize, and mash up datasets.

There is no formal or agreed upon definition of open government data. As the Open Knowledge Foundation defines it, "A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike." For open data provided by government, the defining characteristics fall into two general buckets: legal and technical.

Open Government Data Requirements		
	Legal	Technical
Accessible	No intellectual property or other restrictions on use	No paywall or encryption
Computable		Standardized, non-proprietary formats and bulk downloads

Figure 3 – Open Government Data Requirements

Legally, governments and data owners must not use copyright, patent, contract, or other law to prohibit access to or reuse of open government data. Public sector datasets are also usually presumed to be free on the grounds that the taxpayer has already paid to collect them. [43] Similarly, open government data should not be subject to technical restrictions, such as allowing access only to those who have specific permission to use it.

Overall, the data have to be reusable in any way that the user decides. This means the data must be: 1) published in a standard format; 2) published without proprietary conditions; and 3) available online as a bulk download rather than only through single-entry lookup. When data appear in standard formats, like the one used by Excel and other spreadsheets known as "comma separated values" (.csv), then most statistical or graphical tools can easily ingest and read the data. Neither paper forms nor digital image files are technically usable for the purpose of doing aggregate analytics until the data are extracted into computable formats. In addition, data in certain proprietary formats cannot be understood or processed without purchasing special software or hardware. If, for example, a Microsoft package is required that doesn't work on a Linux or Mac operating system, the release does not meet the criteria for open government data. Finally, without the ability to bulk download the files in a dataset, a third party would have to manually input one data element at a time to analyze the data — a completely impractical approach.

The diagram below shows how several different kinds of government data fit on a spectrum the authors describe as running from "inert data" to "adaptable data." Most of the examples they give of "adaptable data" — those above the horizontal line — fit the description of open government data.
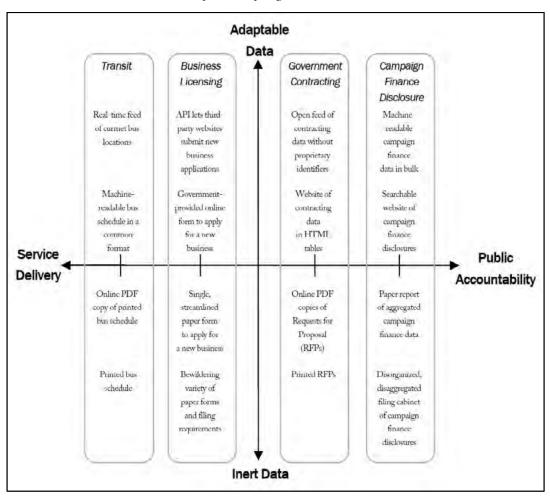

Figure 4 – Open Data Framework
Harlan Yu and David G. Robinson, The New Ambiguity of 'Open Government,' 59 UCLA L. Rev. Disc. 178 (2012).

Why 990 Data?

Highlights

- Nonprofit tax returns can help bring transparency to the nonprofit sector because their contents go far beyond basic finances. They also detail the mission, activities, and organizational and governance structure of each nonprofit.

- While 990 data may be the best starting point for understanding the nonprofit sector, the data quality can be improved. Working with the data in an easily usable form will tell us where improvements are needed and could drive an interest in better data collection and quality control.

With a few exceptions for certain kinds of tax-exempt institutions such as churches, [44] nonprofit organizations in the United States are required to file one of the IRS 990 tax forms. [45] In fact, they automatically lose their exempt status if they fail to file at least once every three years. (For more detail about the requirements, questions, versions, and history associated with the Form 990, see both Appendix 2: The Federal Forms.

	Filing Requirements	Types of Entities That File	Alternatives for Filers
Form 990	Required for organizations with gross receipts equal to or greater than $200,000 or total assets equal to or greater than $500,000, and certain other types of organizations.	Most exempt organizations other than a private foundation or tax-exempt charitable trust may file a Form 990.	Organizations that fall under the Form 990 requirements have no alternative. However, Form 990-N and Form 990-EZ filers may file the long Form 990.
Form 990-PF	Must be a "private foundation" or non-exempt charitable trust.	Only private foundations and non-exempt charitable trusts may file.	Private foundations and non-exempt charitable trusts may only file the Form 990-PF.
Form 990-EZ	Generally required for organizations with gross receipts of $50,000 or more who are not required to file the Form 990 itself.	Medium-sized nonprofits.	Nonprofits that fit the 990-EZ requirements may choose to file the Form 990 instead.
Form 990-N	Only organizations with gross receipts of $50,000 or less.	Small nonprofits.	Nonprofits that fit the 990-N requirements may choose to file the Form 990 or 990-EZ instead.

Figure 5 – What Entities File the Form 990?

The long version contains twelve pages of questions whose answers fill about 1,000 fields with check marks, numbers, or text. These questions cover much more than just income and expenditures. Because the IRS relies on the Form 990 not only to check on taxes owed, but also to verify that the filer is operating in a manner consistent with tax-exempt status, the form was significantly revised in the 2008 tax year to ask nonprofits to describe their missions, their organizational and governance structures, and the services they provide in more standardized way.

The new form increased the number of standardized attachments for certain filers and included new questions on government, management, and disclosure.

The 990 reveals more about a nonprofit's operations than comparable tax forms do about individuals or private companies. As IRS instructions for Form 990 explain, "Some members of the public rely on Form 990 or Form 990-EZ as their primary or sole source of information about a particular organization. How the public perceives an organization in such cases can be determined by information presented on its return. Therefore, the return must be complete, accurate, and fully describe the organization's programs and accomplishments." [46]

Although we usually speak of the Form 990 generically, there are several different forms within the 990 series of returns. Some can be filed either electronically or on paper while others must be filed in one format or the other, resulting in a hodgepodge of formats and forms. Organizations that file the Form 990 and file "at least 250 returns in a calendar year, including income, excise, employment tax and information returns," [47] and have $10 million or more in assets are required to file electronically (private foundations with at least 250 returns must file electronically regardless of asset size). [48] Private foundations file the Form 990-PF, which includes a list of grants made during the year. Nonprofits with gross receipts of more than $50,000 must either file the Form 990 or the 990-EZ, depending on their levels of revenues and assets; those with receipts of less than $50,000 must file either the complete 990, 990-EZ, or a new electronic 990-N, also known as the "e-Postcard." Nonprofits with unrelated business income file a 990-T as well to report and pay tax on unrelated business income.

The core Form 990 requests information on the following:

1. Basic facts about the filer

2. What kind of programs does the filer run and how much does it spend on them?

3. Who are the filer's board members and how does the filer govern itself?

4. Did the filer change in any significant way during the year?

5. How much income did the filer receive and from what sources?

6. What did the filer spend its money on?

7. How much do top earners get paid and salary information

8. Did the filer engage in any insider or excess benefit transactions during the year?

9. Basis for public charity status (if a charity)

10. Does the filer lobby or exercise the 501(h) election? [49]

The Life of the 990

Highlights

- As with many government datasets, there are legal (including statutory) technical, and cultural impediments to openness baked into the process of collecting, processing, and publishing 990 data.

- While e-filing is an option, it is underutilized in collecting 990 data.

- Models for electronic filing from other agencies, such as the SEC, have not been adopted for the Forms 990.

- The IRS is contemplating an initial release of 990 Master File data in ASCII preceded by a project to assess the quality of the data. That undertaking remains a research project. IRS staff in the Statistics of Income (SOI) division are analyzing 990 data. The outcome of the project is yet to be determined.

In order to understand both why 990 data aren't truly open and how they can be made more so in the future, it is important to understand the path the 990 follows from preparer (or authorized representative) to the public. These steps include data submission, intake, review, and release. At each stage in the process, there are impediments to openness built into the process.

Barriers to Efficient and Timely Filing

In order to maintain their federal tax-exempt status, nonprofits are required to prepare and file one of the Forms 990 each year. They must file by May 15 if their fiscal year coincides with the calendar year, but may file up to two extensions, which pushes the deadline back to August 15 and November 15, respectively. Many organizations take advantage of the extensions, thereby delaying the public availability of their filings. The desire for additional time creates an incentive for paper filing, since mailing an incomplete paper form allows an organization to claim timely compliance. The IRS informs the organization of the missing information and will go back and forth with the filer until the form is complete. When submitting an electronic form, the nonprofit is notified immediately if the form has missing or inconsistent information, and the nonprofit is not able to file until the form is rendered acceptable. So electronic filing requires the nonprofit to have all the information at the time of filing, unlike the paper filing method.

Thus, e-filing cuts down on gross errors since electronically submitted forms are machine-checked for completeness and consistency before being accepted. Nonetheless, the IRS does not have the regulatory authority to require universal e-filing without some action by Congress. [50] Only the largest and smallest nonprofits are required to file electronically. Most nonprofits have the option to file electronically or on paper, and the vast majority prefer the latter. The section on "The Prospects for E-Filing," later, discusses these issues in more detail.

Even when a form is e-filed, it is not processed in a way that makes full use of the structured data it contains. The IRS eventually converts each return from an XML (Extensible Markup Language) extract to an image file. Thereafter, the return is processed just as if it had been scanned from paper [51] to ensure equitable treatment of the two classes of filers. As long as e-filing is optional rather than mandatory, the agency is likely to remain dedicated to treating paper and electronic filers identically so as to eliminate any perceived inequity between the two types of filers. Bob Ottenhoff, former President and CEO of GuideStar, doesn't think this will change: "This is an equity issue for [the IRS] and it's a public policy issue. They've got a lot of rules on what they'll do. It's great when applied to what they'll do to the individual tax returns, but it complicates the world in the nonprofit information sector." [52]

Paper and Electronic: Two Paths, Two Processes

Paper and electronically filed 990 forms are initially processed through two different paths by two different divisions of the IRS.

Paper forms go through a complex process overseen by the Tax Exempt and Governmental Entities (TEGE) division diagrammed below. First, the Forms 990 are sorted out from other tax forms at the IRS processing center in Ogden, Utah. Any money included with the forms is removed and deposited. The forms are then batched, put through a tracking system, and sent to the imaging department for scanning and conversion into image files. Forms 990, 990-PF, and 990-EZ that are filed on paper and then batched and imaged by the IRS are stored in a database hosted by IRS Research Analysis and Statistics (RAS). The internal-use-only (un-redacted) images are used in research and certain data processing. External-use images undergo a redaction process before they are made available to the public.

Data from paper returns — not images — are manually transcribed as a part of IRS processing. Data from paper returns are not part of the Modernized e-file System (MeF).

The scanned images are used both for research within the IRS and for external publication to comply with statutory transparency requirements. The IRS Research and Statistics (RAS) database includes images of the forms as they were received. The RAS office in IRS maintains a database of imaged paper returns. Forms 990, 990-PF,

and 990-EZ that are filed on paper are imaged by the IRS. The images are stored in a database hosted by the IRS RAS office. The internal-use-only (un-redacted) images are used for data analysis, research, and the development of datasets made available to the public on www.IRS.gov under TaxStats. External-use images undergo a redaction process before they are made available to the public. Data from paper returns—not images—are manually transcribed.

As Figure 6 shows, electronically filed returns begin on a different path but end up going through the same process as paper files. Forms 990, 990-PF and 990-EZ that are filed electronically are subject to certain checks for "completeness" when they are received by the IRS. These returns are immediately reviewed for obvious errors and are not considered as filed until corrected. Once they have passed that step, XML data from these returns are extracted and used to populate the MeF database. The XML data are also used to create facsimile images of the returns for the same internal and external uses as paper filings. SOI staff use data from the Returns Transaction File (both paper and electronic filings) to select, on a weekly basic, the stratified sample for the 990-series studies.

The sample data, whether electronically filed or filed on paper, are still converted into PDF format which requires the use of Adobe Acrobat but which allows IRS officials to view the data in their systems. At the same time, the data from e-filed returns are put into the scanning process to create image files that can be made public just as images from paper returns are, thus converting machine-readable XML data into non-computable images.

Currently, image files of nonprofit returns provide the only source of exhaustive data on the corpus of Forms 990 filers. Efforts have been made to provide certain data points from the Forms 990 to the public. Research staff in the IRS Statistics of Income division are undertaking a research project to study release of Master File data on nonprofits to the public in a downloadable database form, using the basic computer language ASCII. (Even though electronic returns are filed in XML, a more sophisticated computer language; it appears that they will be converted into ASCII for the public.) At this time, the project remains a pure research effort. Future decisions about those data are yet to be made and will depend on the outcome of the project. The file they are examining is a very large extract with upwards of 500,000 returns across three form types: 990s, 990EZs, and 990PFs from 2011. The data being researched will include 400 fields from those returns.

Since the IRS does not necessarily review all data in the Master File for accuracy, there is some concern about making data from the Master file more broadly accessible. Hence before releasing the data publicly, the IRS plans in Spring 2013 to compare the Master File extract against a carefully vetted and corrected sample from the Statistics of Income division in order to assess data quality prior to wider dissemination.
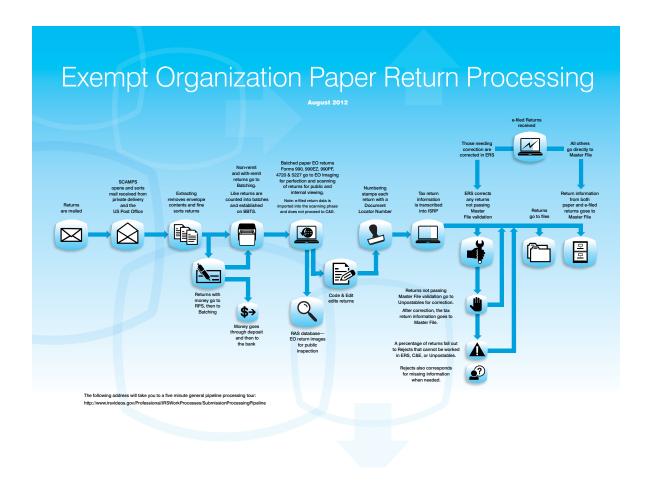
Figure 6 – Source: IRS Exempt Organizations Division

The Publication Process: Transparent, But Not Truly Open

By law, 990 returns must be made available to the public. According to the Internal Revenue Code 6104(d), an organization must provide copies of its three most recent Forms 990 to anyone who requests them, whether in person, by mail, fax, or e-mail. Additionally, requests for a specific return may be made to the IRS using Form 4506-A.

The IRS publishes all the Form 990 scans as TIFF [53] image files on DVDs (after redacting each Schedule B to protect the names of individual donors). [54] Because many nonprofits file extensions, the IRS always has tens of thousands of returns to process and publish each month. Sometimes over 100,000 filings are released at once, filling more than 20 DVDs. It is worth noting that TIFF format (unlike a PDF) treats each page as its own file, meaning a 30 page document is 30 separate TIFF files, leading to voluminous numbers of files on each DVD.

These DVDs contain image files that can be read by humans but not processed by computers. PDF files would, at least, be capable of being indexed and searched by Google in a way that image files without accompanying metadata cannot be. The key point, though, is that the IRS does not distribute the entirety of the Form 990 as truly open data — that is, as a downloadable data set covering every variable within the Form 990 that can be used to create new visualizations or to perform analyses. Data are made available in other ways and as a part of sample releases.

For example, a sample of 990 information that includes the weights necessary for expansion of the sample to population-level aggregates, allowing valid estimates of the population of returns, is available to the public in machine-readable form. [55] The 990-series sample employs a "stratified random sampling" technique. In other words, for sampling purposes, the 990-series populations are classified into like groups (or "strata")—in this case,

based on the amount of assets. The likelihood of a 990 return in the full population being selected for the annual sample is determined by its stratum. For the sample, the largest organizations — nonprofit entities with assets over $50 million and private foundations with assets over $10 million — are selected at a rate of 100%. Lower asset-based strata are sampled at lower rates. As a result, most of the largest organizations are in the sample year after year.

This sample prepared by the Statistics of Income Division is designed to capture the financial activity of the nonprofit sector with the highest feasible level of precision. One primary purpose of the resulting datasets is research and estimation work undertaken at the IRS, the Department of Treasury, and the Congress. Those datasets are also made available to the public in machine-readable form. Research work is undertaken in the private sector using those datasets.

But these releases are both not widely known about and fall short of covering the entirety of either the form or the corpus of 990 filers. This is more than a technical difference; it makes a huge difference in the usability of the data.

To illustrate the difference between digitized image files and open data, imagine that you are trying to fill an important position in your organization. Instead of having candidates' files only on paper, you may find it helpful to store their resumes on a computer for easier retrieval and more secure storage. But think how much more useful it would be if, in addition to PDFs of resumes, the information you care most about — including the candidates' names, degrees, grade point averages, years of experience, and other attributes — was compiled into a spreadsheet, too. With a spreadsheet, you could then sort, search, share, calculate statistics, and perform other computations to help identify the perfect candidate.

Several government agencies have gone to considerable lengths to transform their image files into truly open data. For example, the National Archives and Records Administration (NARA) recently released the images of the original handwritten records of the 1940 Census labeled by address and began to turn those records into usable data. In order to compile a dataset suitable for searching and studying the information locked within these forms, NARA has entered into agreements with for-profit organizations both to extract the information from the handwritten records and, after three years of their exclusive use, to make the extracted data available as a free and public resource. As a public good, these data provide additional continuity to the historical data provided by the Census. These data will allow researchers to use the Census more robustly when analyzing historical data sources.

Similarly, the United States Patent and Trademark Office (USPTO) entered into an agreement with Google to extract raw data from millions of digital image files of patent applications that were previously viewable only one at a time. In contrast to the National Archives deal, Google immediately made the data searchable in a new search engine (google.com/patents) and also freely downloadable as a computable and shared public dataset.

Finally, the Securities and Exchange Commission (SEC) has overhauled its data collection practices to provide open data. The SEC used to accept filings in any electronic format and then make them available to the public one at a time upon request. But machines could not easily identify or extract the data they contained. Some private companies made a business of manually retyping the information locked within these individual files so they could sell the resulting dataset. In addition to other customers, the SEC itself purchased its own data back in structured, bulk, and computable form. The Commission thus used public funds to pay for this public data twice: First to collect it, then to buy it in a form that they and others could use. By 2009, the inefficiency of this practice drove the SEC to begin requiring that filings be submitted in XBRL, which stands for eXtensible Business Reporting Language. Because XBRL is a standard, open, and globally accepted format, computers can easily ingest SEC filings for reuse.

Making 990 data computable requires either changing how the data are collected a priori, as the SEC did by switching to an XBRL filing requirement, or extracting and converting the data from the image files into a re-usable form as the USPTO did. Compiling such a dataset ex post requires either retyping the entries by hand from each form, or scanning the forms through an "optical character recognition" (OCR) system, which converts the digital images into machine-readable data. Currently, the IRS purchases back data from GuideStar that is extracted ex post from the TIFF files that the IRS releases. [56] To be effective, OCR often requires a custom-built algorithm to convert the file.

These methods are not completely foolproof in any case, and each requires extensive checking to ensure accuracy. To make matters worse, the optical character recognition system now used by the IRS was designed for an old version of the Form 990, and was not updated when the form was changed in 2008. A relatively modest investment in new OCR software could have a significant impact on the usability of Form 990 data fairly quickly.

OPEN DATA CASE STUDY: THE UNITED STATES PATENT AND TRADEMARK OFFICE: WHAT ONE AGENCY DID WITH IMAGE FILES AND NO BUDGET

Each year inventors file upwards of 400,000 patent applications with the United States Patent and Trademark Office, which only recently brought down its backlog of waiting applications from a million to half a million. By law, the USPTO publishes patent applications eighteen months after filing. Those applications describe some of the most path breaking (and some of the most frivolous) advances in modern science ranging from life enhancing drugs like Lipitor to the application for the method of exercising a cat with a laser pointer.

Analyzing these applications could yield tremendous insight about the state of innovation in our economy. Researchers could assess the scope and quality of patent claims using the latest statistical and natural language processing techniques. Inventors could quickly learn whether an idea like theirs had already been patented, and might be inspired to develop more innovative products. And investors could mash up the data with SEC filings to get better insight into which companies are most innovative and best run.

Until 2009, economists, journalists, inventors, and investors could only look up one patent at a time on the USPTO website. Each patent application was stored as an image file rather than as part of a searchable database. The servers at the USPTO were so old and overburdened by the work of managing millions of pages of documents and attachments that the office has to throttle access to prevent the agency from crashing altogether. In 2010, to make the data usable and accessible, including for itself, the USPTO put out an RFP for a no-cost contract to extract the data from the image files and make it searchable and downloadable.

Google took on the challenge. The company's executives decided that the benefit of creating a more transparent and data-driven patent system would justify the laborious process of extracting the data from image files. Over the course of the year, they used optical character recognition technology combined with manual retyping of data to make the entire historical corpus of patent data available. Whereas Google had launched a rudimentary patent search engine as early as 2006, the USPTO release of the corpus of patent data transformed a partial into a comprehensive and reliable resource.

This private sector search engine is so fast and accurate that it has supplanted most uses of the USPTO search tools, alleviating the burden on the agency and providing better data to the public at the same time. Google also makes the data downloadable and reports that between 1-15 terabytes of patent data are downloaded each day. One terabyte is equal to sixty stacks of typed paper as tall as the Eiffel tower. Presumably people are downloading patent data to start patent-based businesses, do research to inform their own R&D, and research the state of innovation.

Figure 7 – USPTO Case Study

WHO USES THE 990S NOW?

HIGHLIGHTS

- A handful of organizations go to the trouble and expense of extracting 990 data on their own, including the IRS Statistics of Income Division.

- Third parties cover the cost of these efforts through a combination of foundation funding and cost recovery through fees paid by researchers, media, businesses, and others.

- The resulting data are neither open nor free to the public, and thus not readily available to a wider community to develop innovations.

Publicly available data are not always available for free. Members of the public must pay thousands of dollars for DVDs that contain the Forms 990 filed with the IRS. But these same DVDs are available free of charge to some third parties that may help make at least some of the data available to the public. For example, the media and other government agencies can order the DVDs at no cost. (For more details about who uses 990 data after their release by the IRS, see Appendix 3: Current Sources of Form 990 Data.)

A number of entities that have a longstanding relationship with the IRS, including GuideStar and the Foundation Center, obtain DVDs regularly from the IRS. [57] GuideStar has approximately 5 million Form 990's. The company goes to great lengths and expense to extract data from image files. Though rare, some Form 990's can run over a thousand pages. The extracted data are so valuable that GuideStar sells it back to the IRS for its own research for half a million dollars annually. [58] The Foundation Center offers its own classification scheme and special services, including a database of grant information from over 100,000 private foundations drawn on data from the 990s. For an average shipment of Form 990s, the Foundation Center requires approximately two weeks to convert and post files in PDF format. [59]

The National Center for Charitable Statistics (NCCS) at the Urban Institute runs checks on what the IRS maintains in its raw Return Transaction File (RTF) database by cross checking the cleaned Statistics of Income Division data sample against the published DVDs. It fixes the inconsistencies where it can, adds some standard classification codes, and posts the results on its website. All this cleaning, verifying, and supplementing typically starts at the end of January when the RTF file is released and takes at least three months. Some statistics computed from the resulting "core" database are available for free. Both GuideStar and NCCS also charge fees, often at negotiable rates, for bulk downloads, for special analyses, or for rekeying data fields of special interest that are not among the core variables.

Many other websites and institutions also work with 990 data and other information about the nonprofit sector, some of whom make this information searchable to the public. These include GiveSmart, Charity Navigator, Givewell, Charity Blossom, the Johns Hopkins Center for Civil Society Studies, the Center on Philanthropy at Indiana University, the Center for Effective Philanthropy, Grantmakers for Effective Organizations, the National Center for Family Philanthropy, the Philanthropic Initiative, and the Philanthropy Roundtable. (For more on the major organizations working with Form 990 data, see Appendix 3.)

Much of the funding for compiling, cleaning, and assembling these data platforms comes through grants from large foundations. The Foundation Center raises and spends approximately $600,000 annually for extracting data from scanned images. The NCCS spends $40,000 to key the data from the scanned files and an additional $40,000 to $60,000 to manage the collection process and verify the Return Transaction File. GuideStar spends about $1,000,000 annually in extraction costs. To be comprehensive, this involves manually typing in those 990 forms that are paper-filed, including some that are hand written. And even with these expenditures, these organizations extract only a portion of the data from each Form 990 and do not extract data from all filers.

For the past twenty-five years, the IRS itself has also collected microdata from a sample of Forms 990, 990-PF, and 990-EZ filed. The work is done in the Statistics of Income Division. Data from the returns in the sample are subject to extensive testing, research, and re-allocation. Currently, over 5,000 data fields from a sample of 25,000 nonprofit organizations are included in those microdata files. The microdata files, which include several thousand data fields from Forms 990 and 990-PF, comprise the population of the largest nonprofit charities (about 7,500) and the largest private foundations (about 6,000). These machine-readable files are made available to the public, at no cost, on www.IRS.gov, under TaxStats. In addition to downloading the complete microdata files, TaxStats users can access the SOI Bulletin and annual and time-series tables based on the Form 990-series sample through the web pages. [60]

As previously mentioned, the IRS does collect some data in an electronic format using the Modernized e-File system (MeF). If the IRS made the e-filed data available in open form, the Urban Institute estimates that it would save $350,000 in the cost of data conversion and $250,000 in savings from a reduced need to conduct quality assurance checks. These data represent only those organizations that currently file electronically, either voluntarily or as required by law. However, if every filer was required to file electronically and their information was made available in computable form, then it is estimated that the private sector and IRS would save approximately $7.4 million dollars total. [61]

Current practices for making 990 data available have clearly not kept pace with technology. The status quo works well enough when the IRS is asked to supply Forms 990 one at a time, but is not organized to provide computable, comprehensive, and freely accessible bulk datasets. There is a widely recognized need to modernize the process by improving the form in which government 990 data are delivered, by making it easier for third parties to convert 990 data into usable forms, or both. In fact, nearly everyone we consulted is expecting 990 data to become open and computable eventually, and they are formulating both business and public service plans on that basis.

Why Liberate 990 Data?

Highlights

- This section illustrates the potentially transformative impact of open 990 data through a series of examples, also known as "use cases."

- In weighing the costs of data liberation, it is important to identify the likely users of the data and those who will benefit from data re-use.

Open data about nonprofits could benefit a variety of people and groups. A larger group of researchers, technologists, academics, and others could use 990 data to deepen their knowledge of the nonprofit sector. Donors could get better information about how closely an organization's actions fulfill its main objective. Nonprofits could learn how other nonprofits are run, information they can use to improve their management practices. Job seekers and for-profits that provide services to nonprofits would also benefit.

Comparative data about nonprofits' operations and impact would be especially valuable to support public and private policy decisions. [62] A computable database built from Forms 990s could not only yield insight about nonprofits' missions, grants, capabilities, and governance, but also provide a framework for organizing and comparing supplementary information not collected by the IRS. All this would make thinking about nonprofit capital markets more natural and more efficient.

While large databases about nonprofits are now available from GuideStar, the Foundation Center, NCCS, and other sources, these institutions need to charge for access to their databases and limit their use in order to recover the costs of converting government data into something more usable.

And because these organizations are forced to work with IRS data in an inefficient form, their databases are not as current, accurate, or high-quality as they could ideally be. [63] The examples below show what will become possible if and when 990 data are widely available as open data.

Illustrative Uses for Open 990 Data

Here are a dozen "use cases" that illustrate in more detail what might be done with access to computable data from Forms 990s (often in combination with other data). While some of these goals can be accomplished now, more readily available, free, and current open data could make the same work less cumbersome, costly, and clunky. Such a dataset, together with suitable input and output standards through API's (application programming interfaces), would bring much more liquidity, agility, connectivity, and interoperability to information and to calculations concerning the nonprofit sector. Based on what has happened following the liberation of other government data, there are probably many more uses that no one can imagine until the information is in hand—including, in particular, new and rewarding ways of earning money for those who know the sector well.

Locate expertise and resources.

Imagine walking down the street, seeing some homeless people, and using your Smartphone to map out which nonprofits sponsor shelters in this particular neighborhood so you can volunteer. The website Catchafire offers a matching service to connect volunteers to nonprofits, but the organization has to solicit these opportunities manually now. With open nonprofit data, it could use 990s to connect volunteers in a community or online to an organization that matches their interests. Ditto disaster planning, which necessitates identifying sources of social services quickly. And with better 990 data, nonprofits could identify experts who have worked with similar organizations to consult and hire.

Aggregate and compare programmatic expenditure trends.

If open, Form 990 data could also be linked to information in other datasets. For example, the NCCS combines Form 990 data with Census Bureau data in its Community Platform. It might be particularly useful to track how programs of a given type or in a given region respond to economic cycles. Johns Hopkins University's Center for Civil Society Studies does extensive work on tracking the economics of the nonprofit sector. With open data their economic analysis could be significantly improved or, at least, conducted with greater ease and efficiency. [64]

Identify leaders and their connections.

In addition to graphing influence networks as groups, like the Center for Responsive Politics typically does, knowing who runs a nonprofit can also help detect fraud. Hugh Jones, Supervising Deputy Attorney General at the Hawaii Office of Attorneys General, informed us that Attorneys General have occasionally found the same person collecting full time salaries from several different nonprofits. Such discoveries are now made by happenstance. With open data, they could be detected through computer analysis. [65] Similarly, Form 990 requires charities to disclose loans to or from current and former officers. Transactions like these are illegal in some states, and could be discovered or monitored more easily if the data were readily available and searchable.

Aggregate and compare specific income trends.

Knowing more about where certain streams of nonprofit revenue come from could help inform debates about changing the rules for charitable deductions, for example. The USASpending dataset of national government expenditures is open and downloadable, and when mashed up with 990 data could provide revealing insights into whether a "smaller" government is really smaller or simply delegating spending to non-governmental organizations, for example. Code for America — an analog of Teach for America that sends young, tech-savvy innovators to work on local innovation in municipalities — would also be interested in 990 data to help them develop apps that showcase the nonprofits serving particular communities. [66] In other words, they want to use the data to build consumer-friendly tools for visualizing who delivers what service.

Promote transparency and trust.

The 990 should be public because public oversight is an important and integral part of our system, reported one official at the Department of Treasury. The whole of charitable deduction rests on the assumption that individual tax payers can make wise decisions about how to best address the charitable issues in their own communities.

Concerns raised about fundraising and functional expenditures by nonprofits, for example, cannot all be addressed by calculating simple ratios. Enabling more sophisticated analysis could help distinguish good and bad outliers.

The Navy Veteran's Association scandal [67] demonstrated how hard it could be to detect false claims without an

aggregate database that can cross-check 990 filings in different states. In that scandal, a con artist set up a fake Navy Veterans' charity and collected millions of dollars before he was caught. A database that allows for cross-checking Forms 990 filings across various states would allow regulators to more easily track and investigate organizations that may falsify information on their activities or officers.

Assess financial soundness.

Policy makers may want to know how well different parts of the nonprofit sector are doing, and donors may want to know how well different potential recipients are doing relative to comparable parts of the nonprofit sector. With enough data, models can be devised and tested to predict insolvency, which is how GuideStar approached a recent "Data Dive" aimed at creating tools for the sector. [68] Givewell, a charity evaluator that helps donors assess charities for their financial soundness, would be dramatically aided by bulk data downloads.

Track and compare labor statistics.

The demand for basic information about nonprofit salaries is strong enough to support commercial suppliers who compile and sell it. [69] With some states proposing legislation to regulate such salaries, access to accurate analysis and statistics can only become even more important. Employment within the nonprofit sector is another important macroeconomic variable that can be deduced by aggregating Form 990 data to use in conjunction with data from the Bureau of Labor Statistics (BLS). The Johns Hopkins University Nonprofit Economic Data Project tracks changes in employment, wages and finances over time and in comparison to other industries using BLS data with 990 data that must now be manually and painstakingly processed. [70] (The Project is a collaboration between the Center for Civil Society Studies, State Employment Agencies and the U.S. Bureau of Labor Statistics, and State Nonprofit Associations.) Given the importance of the nonprofit sector to the country's overall employment rates — estimates suggest that the nonprofit sector accounts for 10% of the workforce in the U.S. — more useful data about nonprofit employment is especially needed.

Check investment and program alignment.

Venture philanthropy is a growing trend in the nonprofit sector. A foundation dedicated to genetic research on a given disease might use better 990 data to search for potential partner foundations that invest in or hold biotech companies in their endowment portfolios. Foundations such as the Michael J. Fox Foundation for Parkinson's Research provide grants for firms thinking innovatively about medical research. [71] For such foundations, information about the extent to which investments and portfolio holdings mirror mission statements could be presented in a more accessible manner for regulators and donors.

Aid enforcement and regulation.

Steven T. Miller, deputy IRS Commissioner, testified before Congress that the agency was "somewhat understaffed" in overseeing nonprofits, especially ensuring that they pay taxes on unrelated business income. [72] Better, more usable data would make enforcement more efficient and effective. For example, checking the box on the Form 990 that asks whether loans have been made to directors or officers should constitute a "red flag" for the Attorney General in any state where that practice is prohibited. Without ready access to specific information about suspicious activity like this, there is a tendency to regulate in broad and heavy-handed strokes based mainly on anecdotes. With a Form 990 database, identifying potential fraud could be crowdsourced, giving the IRS and states' Attorneys General "more eyes and ears." As Hugh Jones put it: "Making 990 forms public enlarges the volunteer staff of the IRS."

<div style="border:1px solid black">

GOING ELECTRONIC

For approximately fifteen years, New Mexico did not enforce its nonprofit state registration statutory scheme. Organizations were required to register at the state level, but enforcement was virtually non-existent. The state decided to go from a paper-based filing system to an electronic filing system to improve enforcement. Lawmakers recognized that this change would require adjustment in filing practices and offered a year of amnesty regarding late registration fees to assist with the transition.

During the rollout, filing organizations experienced confusion and inconvenience. Elizabeth Korsmo, New Mexico's assistant attorney general, was one of the point people for the rollout and fielded a number of queries and complaints from the nonprofit community. After receiving advice from attorneys general in Colorado and Hawaii, states with electronic filing systems, Ms. Korsmo enforced the electronic filing requirement with a near zero-tolerance policy. Exceptions were only made when there was some form of extreme hardship, usually related to a physical disability. Otherwise, nonprofit filers were expected to register electronically, even if they had to find a resource center, such as a library, to do so. Since then, fewer than ten organizations out of approximately 5,000 organizations were have been allowed to file by physical means.

There has been some duplication between the New Mexico registration process and the filing of the Form 990. This is an unfortunate necessity since much of the information locked in the Forms 990 is necessary for attorneys general. However, the result is that nonprofit data for the state has become highly searchable: board members and accountants can be searched for across organizations, and comparisons can be made between a nonprofit's current mission statement and any prior mission statement placed on the forms.

</div>

Figure 8 – New Mexico's E-Filing Law

Monitor Lobbying .

Certain regulations restrict how much lobbying and political activity a nonprofit engages in, depending on the type of nonprofit. The Citizens United [73] decision obviously has brought political issues to the fore. The Center for Responsive Politics (CRP) tracks money in U.S. politics and its effects on public policy. CRP manually scrapes a relatively small number of 990s for politically active nonprofits to populate its OpenSecrets website, which tracks which organizations give to whom. "A public source for 990s data," says Executive Director for the Center for Responsive Politics, Sheila Krumholz, "would be a godsend." [74]

Facilitate data "mashups."

Data from totally different sources can often be combined to great effect. Form 990 databases could be combined with mapping data, for example, or with programmatic results data that could help nonprofits gauge their effectiveness in comparison to other similar organizations. Like the current initiatives to create a global system of "legal entity identifiers," [75] an initiative specifically aimed at numbering nonprofits would enhance such capabilities, as could state efforts to standardize their nonprofit registration procedures. An effort specifically for registering nonprofits is being developed by a foundation-funded consortium under the name Basic Registry of Uniquely Identified Global Entities (BRIDGE).

<div style="border:1px solid black">

"Very little is known about [nonprofit and endowment investments], but there is a great deal of excitement on impact investing. It's important to have a better understanding of where the $600 billion dollars nonprofits hold actually goes."

-Lucy Bernholz, Visiting Scholar at the Center for Philanthropy and Civil Society at Stanford University

</div>

Facilitate academic and business research.

Establishing various baseline values and setting up longitudinal studies may now be especially important given the changes taking place both in and around the nonprofit sector. Statistical frames for constructing representative survey samples, for example, depend on having comprehensive data about the nonprofit sector as a whole. More specifically, the Urban Institute could use computable 990 data to pursue its interest in understanding the mix of nonprofit institutions that exist in low-income neighborhoods; the relationships between the level and mix of neighborhood-level nonprofits; and conditions and trends in those neighborhoods. In fact, the academic study of the nonprofits has become so important that Indiana University is building on the research experience of its Center on Philanthropy to launch the nation's first School of Philanthropy.

Mechanics of Creating an Open 990 Data Platform

Highlights

- Before developing any innovative applications with the data, 990s will either have to be collected in or converted into open formats. Other data cleaning and redaction will also need to be performed.

- The cost of freeing data must be weighed in addition to other evaluative criteria for determining which strategy to adopt.

Technical Issues

Creating an open, shared, usable 990 dataset for all will necessitate changing how the 990 data are collected; how they are published; or how they are handled ex post. Before addressing the process for creating and maintaining such a platform, it is important to clarify what needs to happen technically to get from today's world of image files to tomorrow's open dataset.

All filers could be legally required to submit their forms electronically pursuant to legislative mandate. This would not only necessitate mustering the political will to enact the change. In addition, the IRS would need to reengineer the systems it uses to manage the intake and processing of 990s, which were created in an earlier era before electronic filing was contemplated. The IRS could then set up structured forms so that the output would automatically be in an open format and no conversion would be required.

Alternatively, if the forms continue to be filed on paper, the data must be converted into open formats from the existing image files. As image files, the data have limited use. Images can be looked up one at a time but, as discussed, they don't lend themselves to automated processing or correction. Raw data could be extracted from the digital images using optical character recognition or other forms of data conversion that have been well-honed in other domains. [76] Hard-to-read forms, such as those written by hand, also need to be retyped manually. The retyping is usually done two or three times (what is known as double or triple keying) in order to ensure accuracy. Regardless of how the forms are filed, if the data are not input in a way that is designed to output open data, image files will need to be converted into computable formats.

In addition, Social Security numbers and signatures need to be redacted from the data, which would require statutory authority. This will also require better organization and quality control. Again, these kinds of data cleaning techniques have been long practiced by those who work with 990 data and on other open data projects. But it requires work over an extended period by many careful hands to clean up a complex database like this and make it useful.

The Prospects for E-Filing

While the chances for a legislative e-filing mandate are unclear, it would be a positive step that would make it easier to develop an open data platform for data from nonprofit organizations. The chances of success are dif-

ficult to gauge because few systematic efforts have been made at the federal level to promote universal e-filing by nonprofits. Although it is important to note that e-filing was one of the recommendations of Independent Sector's Panel on the Nonprofit Sector.

Financially, e-filing could eventually save millions of dollars in costs for data collection and processing, particularly in reducing the cost of error detection and correction. An immediate obstacle, however, is that universal e-filing would take time, effort, and funds to set up. In particular, the IRS would probably want to simplify the electronic filing of attachments, which is currently a barrier for many organizations. The IRS currently has not provided an estimate of the relative costs of e-filing and paper filing.

Legal research by analysts at the Urban Institute has identified a single provision of the tax code as the major impediment to requiring e-filing: It is IRC sec. 6011(e)(2), which was intended to prevent the IRS from requiring small businesses to submit quarterly unemployment data on magnetic tape. If this provision were recognized as outdated and eliminated, the IRS could mandate universal electronic filing in the same way that it has mandated e-filing for organizations with $10 million or more in assets and more than 250 returns. Such a move would be supported by several facts, as summarized by the Urban Institute:

- The electronic Form 990-N is already required for half a million of the smallest organizations as well as the largest ones. (The total number of returns filed electronically is shown in the table below.) Despite concerns about the burden and their ability to electronically submit their forms, these small organizations have been doing so successfully for more than four years.

- Free software for electronic filing is available from NCCS for organizations with revenues of less than $100,000. [77]

- In its Modernized Filing (MF) system, the IRS already has much of the technical infrastructure in place for electronic filing.

- In addition to potentially saving money, electronic filing would speed the process of making data available to all stakeholders.

Electronic Filing of Exempt Organization Returns, 2012 Projections

	Form 990	Form 990-EZ	Form 990s & 990-EZs	Form 990-PF	All Forms
Total returns	286,152	301,900	588,052	112,700	700,752
e-filed	118,272	74,800	193,072	29,000	222,072
% e-filed	41.3	24.8	32.8	25.7	31.7

Figure 9 – Source: IRS Publication 6049 (2012)

While electronic filing is being debated at the federal level, several states are beginning to implement similar initiatives. Under the auspices of the National Association of State Charity Officials, twelve states including California, New York, and Illinois are likely to participate in an e-filing pilot project with Urban Institute and Columbia Law School's Charities Oversight and Regulation Project. In this pilot, they will provide a single website where nonprofit organizations can go to complete their Forms 990 along with any of the state forms required. Returns can be submitted simultaneously to both the IRS and participating states. Under this system, the same returns could easily be used by third parties as well.

In sum, mandated e-filing is worth continued exploration. But while that long-term goal is being pursued, we need to explore other options for making 990 data available, usable, and downloadable.

Pricing Issues

From an economic point of view, freely downloadable datasets are "public goods" in the technical sense of this term. That is, they are both "nonrival" (use by one party does not diminish use by others) and "non-excludable"

(use is available to everyone once it is available to anyone). Classical examples of public goods include lighthouses, parks, research discoveries, and other openly published knowledge. Without mandates or government intervention, robust funding mechanisms for public goods are notoriously difficult to devise. Hence societies tend to systematically undersupply public goods.

There are significant costs to improving current IRS platforms to process, compile, update, support, and supply an increase in electronically filed Forms 990 to make available in an open, machine-readable Form 990 database. As with other examples from the open government movement, providing an exhaustive catalog of accessible, machine-readable nonprofit data is expensive relative to anything that follows.

OTHER EVALUATIVE CRITERIA

Any new 990 data platform should ideally exhibit ten attributes (listed in alphabetical order), including minimizing cost.

1. **Accountable.** There should be a mechanism for responding to suggestions and complaints and, preferably, proactively anticipating issues, such as the accidental inclusion of Social Security numbers in Forms 990.

2. **Adaptable.** Whatever processes, procedures, categories, and formats become instituted now are bound to need changing over time. A platform should be able to cope with evolving expectations and technological capabilities.

3. **Current.** Information for posting may be released sporadically or according to regular cycles. In any case, a platform should be able to update the collective datasets it provides rapidly to incorporate the best and most recent information available and to track new, dated additions to the data.

4. **Efficient.** A platform should constantly seek out and implement ways to make its operations more cost-effective. Though conceivably necessary due to other constraints or considerations, rekeying information or other duplicated efforts are not signs of great efficiency.

5. **Generative.** The platform and the data it produces should inspire innovative uses and expand new "ecosystems" of users who are not now taking full advantage of 990 data. The development of communities that share active interests in the data will help provide continuing feedback, support, ideas, and results.

6. **Harmless.** Providing data collectively that is already available individually should, in principle, do no harm. But given the unique, vibrant, and varied nature of the nonprofit sector in the United States, care should be taken to avoid unanticipated consequences. Is there any harm created by releasing data and making it publicly available?

7. **Official.** Data submitted to and processed by the IRS is official but not necessarily accurate. Mechanisms should be considered, for example, to improve accuracy and/or to inform downstream users about mistakes and corrections. At the same time, knowing what was filed with the IRS and when — regardless of errors — must also be preserved for purposes of legal discovery.

8. **Ready.** How long would it take to establish a proposed platform and for it to begin delivering data? Are the political or other pre-conditions in place that would make success likely and timely?

9. **Sustainable.** Even if its set-up costs are much more than the running costs, a platform needs to have at least one plausible plan for securing long-term funding. A back-up plan would be wise, too.

10. **Trustworthy.** The data provided should be recognizably unbiased and as accurate as possible. It may be desirable to distinguish between verified and unverified information.

Alternative Approaches to Opening 990 Data

To summarize the problem we face: 990 data are publicly available, as required by law, but they are not easily accessible. The data can only be used in limited ways to create visualizations or mash-ups or to do data analysis because most of the data are locked up in image files and not available in machine-readable and computable form. The IRS advertises DVDs of all the 990s (http://www.irs.gov/Charities-&-Non-Profits/Copies-of-Scanned-EO-Returns-Available) from 2005 to present for sale. However, these DVDs contain only images of the 990 forms (TIF files), not data in a usable format.

In addition, the IRS charges significantly for the data and does not release the electronically filed data in an electronically structured form. A year's worth of Forms 990, Forms 990-PF, Forms 990-EZ, and Forms 990-T on DVDs costs $2,580 and a complete set over $20K. [78] The cost to purchase and, more importantly, the cost to process the TIFF files make them cost-prohibitive for many organizations that might want to use the data, and a significant cost for those who do use the data. [79]

There are four major alternative mechanisms for creating a public good 990 platform: 1) Legislative Mandate for E-Filing; 2) IRS Initiative; 3) Third-Party Platform; and 4) A Priori Electronic Filing. We briefly present a description and analysis of each and then recommend a path forward.

Legislative Mandate for E-Filing

In principle, the most straightforward solution would be for Congress to require e-filing by instructing the IRS to collect all Forms 990 electronically and to make a database of all that 990 information publicly accessible (or, at the very least publish forms in the format in which they were filed such that e-filed documents are released in computable formats). In practice, however, nothing like this is ever straightforward.

Congressional representatives naturally hear conflicting messages about legislation like this. Few people are explicit against efficiency or transparency. Some nonprofits may see e-filing as a burden. Others may not be eager to make their work so readily subject to scrutiny, tracking, and comparison. And some nonprofits that object to e-filing may be quite active politically. There are also companies that make money by creating special purpose databases with 990 data, like those that sell salary information, whose business would be hurt by a more open data platform. However, there have been other movements toward requiring electronic filing for tax purposes that have not created great uproar. It is likely that there would be some discussion around any legislative proposal, but the growing chorus within the nonprofit and tech sectors would gain attention as well as those who may not want increased transparency for the sector. For all these reasons, any change would require both strong advocacy and strong Congressional champions.

Currently, most nonprofit organizations have the option to file their returns either in paper or electronically at their discretion. Only 20%-30% of those with the option to file electronically do so. The Urban Institute convened a meeting of accounting firms who file the 990 on behalf of nonprofits to discuss incentives for enhanced, voluntary electronic filing. Their reasons for continuing to file on paper include familiarity, seeing the paper process as "safer" than electronic filing, the greater security in filing paper forms that won't be immediately returned due to incompleteness, and the ease of providing attachments. Private foundations must file information in the form of attachments that currently are difficult to transfer electronically, and previously proposed changes to the electronic Form 990-PF were set aside because of the amount of work required for the revision. As a result, only a small, resource-rich percentage of private foundations file electronically. In conclusion, without a mandate, most of the sector will continue to file on paper.

Pros: A congressional mandate for e-filing would create strong momentum for making 990 data more accessible and usable. Legislation would overcome any potential uncertainty on the part of IRS officials as well as obviate the need and cost for third parties to extract the data after the fact. The money saved could be redirected to analysis of the data instead. By requiring electronic filing, the IRS would also be able to reduce paperwork and shorten the review process. The IRS goes through an onerous process of converting electronic Form 990s into image files for the purpose of treating paper and electronic filers the same. Mandated e-filing would eliminate that extra step and,

presumably, increase the timeliness in making returns publicly available. It is conceivable to phase in electronic filing over time and to work gradually toward a strategy for making electronic data freely available.

Cons: Legislation is hard and fraught with compromise. Getting attention for the need for e-filing could also be difficult. It would require consensus building in the nonprofit sector, which is highly diverse and includes universities and charities, hospitals, and churches with potentially divergent interests. Legislation will require rallying these many stakeholders in a targeted effort to succeed in an arduous and lengthy process. Given the interests in play, some may fight an e-filing mandate. At the very least, without concomitant changes to the private foundation forms, it can be assumed that a contingent of private foundations would object to electronic filing. The prospect of legislation would have to be broached with the nonprofit community and its allies because of the potential burden to them. Politicians will need to be persuaded that this is an issue worth defending. And even successful legislation might be watered down to a degree that does not achieve the goals of transparency and usability.

(For more details, see Appendix 5: Electronic Filing.)

IRS Initiative: Data Digitization and/or Extraction

The IRS could continue to accept both paper and e-filing, and could assume the cost and burden of creating usable 990 data from those sources. To create an incentive to e-file, the IRS might offer a later deadline for e-filed returns and an earlier one for paper-based filings. Where possible, image files could be used to populate a database using optical character recognition (OCR) to extract the data. Hand-written forms that are hard to scan with OCR could be manually re-typed and error-checked by double or triple keying, namely, having the forms retyped twice or three times to ensure accuracy. Alternatively, the IRS could update its data management processes to make electronically filed data available in raw form together with extracted paper-filed data. By changing the intake and publication process, it could dramatically improve the availability of usable, computable data.

Pros: The advantage to doing this within the IRS is, above all, the perception of reliability and authenticity over data extraction conducted by a third party. While the IRS has no statutory authority or resources to correct the data, raw data made available by the IRS will carry a more official imprimatur than data from another source. In addition, since the IRS already has much of the data in electronic form, it could still treat both electronic and paper-based data equally by releasing both at the same time and in the same format. But the IRS would save steps by having to extract data only from the paper-based files.

Cons: In the current economic climate, spending any additional resources on digitizing or extracting data relating to nonprofit entities might be hard to justify without specific direction from Congress. Resources to digitize or extract data have not been appropriated. In fact, interviewees repeatedly stated that the IRS could not undertake such efforts under current circumstances because they would incur new costs.

Moreover, it isn't clear that a government agency is in the best position to do this work cheaply and efficiently. The IRS would have to hire a company to do the work, and procuring the services of an approved vendor (not just anyone can bid for a government contract) would be a time consuming process that could take more than a year. By virtue of its governmental status, the IRS like other agencies would be required to hire workers to rekey hand-written forms at potentially greater cost than a third party, which can off-shore the work, can charge.

In addition, because the IRS has no authority to correct or clean the data, it would be at risk of posting raw data containing social security numbers and other personally identifiable information that should be redacted first. Furthermore, the fact that the IRS is doing the posting might convey the erroneous impression that the data are guaranteed to be accurate.

Third-Party Platform

Either a new or existing third party could take responsibility for compiling and posting a Form 990 database that is comprehensive and computable. Carl Malamud of Public.Resource.org, the same organization originally responsible for spurring the SEC to create the searchable database EDGAR, has been working toward this goal

and has been systematically posting the IRS 990 DVDs online. As a first step, this obviates the need for others to pay the IRS for public data and to wait to receive the DVDs. [80] Second, he has converted the image files into PDFs with proper metadata, making them at least easier to index and search for online. [81] With PDFs, every time someone types the name of a nonprofit into a search engine, they can more easily find the 990s without having to go to GuideStar. Third, he is spearheading a campaign to get large foundations and charities to persuade the IRS to take the data it already requires to be filed electronically, using the Modernized e-File (MeF), and release it as is in a timely fashion. [82]

Established nonprofit data providers, especially GuideStar, the Urban Institute, and the Foundation Center, already do a lot of work to make 990 data available: They have systems in place for processing the TIFF files on the DVDs that the IRS sends them, and could set up and maintain an open and computable database of 990 data. These groups have experience with both internal and external procedures for extracting and checking the numbers, text, and checkmarks from those image files. However, to make 990 data truly open, the scale of such activity would need to be more extensive than anything previously attempted. Much of the work of preparing and assembling data would, in practice, continue to be outsourced to for-profit vendors but under the supervision of nonprofit experts. One possibility would be for these three organizations to collaborate to create a new, open database of 990 data.

If a third party or collaborative group were to make its 990 database available at no charge, as open data requires, then the relatively high initial costs as well as the relatively low continuing costs would need to be recovered. One option would be to rely primarily on sponsorship by foundations willing to offer core support for releasing raw (potentially cleaned up) data. A second is to charge for incremental services, which might include scrubbing for accuracy, or value-added services such as analysis, visualizations, and custom reports that they can distinguish from access to the basic data. Compensating data providers this way might mitigate their costs and encourage innovation without diminishing the value of 990 data as a public good.

It is important to distinguish here between two types of fees. Charging users for basic access to raw data collected from the public pursuant to a legal mandate goes against the principles of open data (on the theory that this information belongs to the public and federal information, in particular is in the public domain under U.S. copyright law). Even though there are costs associated with cleaning, formatting, and classifying these datasets for public use, making these data truly open, meaning publicly available for bulk download and wide use, may expose issues to a wider community whose solutions may be less cost prohibitive. But if a user wants special services, such as an unusual statistical analysis or the compiling of additional information, the platform would have to spend incremental time and money honoring such a request. A price at least high enough to cover those marginal expenses makes sense. Moreover, platforms like NCCS who do perform such services can also insist on posting the results for use by all, increasing the public good they provide. Such an exchange adds value for everyone.

Pros: A third-party strategy obviates the need to undertake a formal procurement. The work could begin as soon as the image files are uploaded. Doing this work outside government also makes it easier to identify information that requires redaction, such as Social Security numbers, since a third party could establish standards of accuracy that the IRS is not empowered to do.

The benefits would be especially great if multiple entities were to collaborate in a consortium-based approach to sharing the work of extracting the data. Given that funders already spend millions on data extraction with multiple entities, this community source approach could represent a significant saving and a faster way to make the data available to all.

The third-party approach opens up more flexible avenues for funding the work, too. Companies that already engage in data extraction, such as Google, might donate their services. Foundations who already pay organizations (repeatedly) to do this work could pay once for the creation of a shared, public data repository. Organizations that do the extraction, such as GuideStar or NCSS these, as noted above, might share what they have because collaboration would bring down the extraction costs for everyone and free up resources to invest in developing innovative, value-added products. University researchers and other groups could apply for grants to extract this data. Those with an interest in a healthy and innovative nonprofit sector could fund a prize to invite unanticipated but willing parties to do the extraction. Any or all of these entities could collaborate by divvying up the work of

extracting, redacting, cleaning and analyzing the data, sharing the fruits of their labor with one another so that the community as a whole has access to the corpus of usable 990 data.

Cons: Having third parties do the data extraction requires coordination or, at the very least, instigation to find ready, able, and willing volunteers with the resources to process this volume of information. Whereas Google took on Patent Office data to fix and monitor a broken intellectual property system, the head of that project emphasizes that there is no similarly compelling business case for 990 data. Given that there aren't assignable profits to be appropriated, no one might step up.

Another downside is the perception of unreliability. Data extracted by companies or other organizations might be perceived as of lesser quality than data provided by the IRS. "The advent of e-filing made it seem obvious that the information would get out more quickly, but there were still a fair number of errors in the data," said Linden Smith, Managing Director of the National Economics and Statistics Tax Sampling team at PricewaterhouseCoopers. "With aggregations of raw, electronic data, errors give you large distortions and deviations. It's critical to think about overall data quality." The other real downside to such a strategy is sustainability — the difficulty of finding the resources to make both past and future 990 data available on an ongoing basis. However, if third parties begin to develop platforms now to make 990 data open and available, either individually or in collaboration, the demonstrated value of the data may open opportunities for funding, create the impetus for Congress to mandate e-filing, or lead to another sustainable path.

A Note on Crowdsourcing

Enlisting an army of online volunteers to help create or contribute to the 990 dataset through retyping returns is an option, though an unlikely long-term solution. The Old Weather project, for example, organizes volunteer teams to transcribe World War I Royal Navy shipping logs to help scientists model climate change projections. Who would have thought that 30,000 people would volunteer? It is therefore not inconceivable that a volunteer army of crowdsourced labor might be enlisted to retype portions of 990 forms to create an open 990 dataset. However, unlike in the Old Weather case where the data are unavoidably hand-written, much of the 990 data has been electronically filed and could be made available in open form. Hence volunteers might be hard to persuade to do what they perceive as make-work. Crowdsourcing, however, could be part of a short-term, pilot project to create some usable data for mashing up.

A Priori Electronic Filing

One alternative to extracting the data is for some of the data to be filed electronically a priori — not with the IRS, but directly with the third party managing the open data project. Suppose that, after e-filing their returns, nonprofits had the default option of also sending their Form 990 e-files to a platform set up to collect, compile, and post them. A simple "plug-in" added to the software already used for e-filing would do. Nonprofits could also simply commit to making their data available, as the major foundations have made in the "Reporting Commitment" as a part of the Glasspockets initiative. [83] Each electronic Form 990 received directly by the "Public Resource/GuideStar/Foundation Center Consortium" would need little further processing. Given the cost of extracting data from image files, it might even be possible to save money by offering nonprofits and their CPAs incentives to e-file both with the IRS and with the third-party provider. This might also motivate the IRS to make its e-filing process able to accept lengthy attachments since some larger nonprofits will persist in filing on paper until that is possible.

Pros. Creating an alternative route to filing 990s such that they can automatically be published in open and computable form has the advantage of cheaply creating a clean and useful source of data. It could be an attractive option to many entities who will want to be known for being more transparent and who, as a matter of principle, will wish to contribute to creating such a shared public resource to the end of improving effectiveness in the sector.

Cons. Such an alternative is limited. It will attract only some filers, likely the largest and smallest, rather than a representative or comprehensive sample. Hence the resulting data may be of limited use. What the industry saves

in the costs of creating the database, it will have to spend on marketing the effort and getting organizations to take advantage of it. Many will worry about participation and giving data to any entity other than the government for fear of misuse and mishandling of their information. It is unclear who needs to participate initially in order to catalyze greater participation in the long run.

Recommendation for Creating Usable 990 Data

We recommend pursuing two paths in parallel. First, we recommend developing a third-party approach to extracting the 990 data in order to demonstrate the value of such transparency. Second, we recommend pursuing parallel legislative approaches, including conversations on the Hill both with those who have demonstrated an interest in the nonprofit sector and with those who are supporters of open data. These discussions should be accompanied by demonstration projects and examples to show how the taxpayer, the IRS, and Congress would benefit from usable 990 data. Additional legislative approaches include state level efforts to include an open data provision in the Model Nonprofit Corporation statutes, and a gradual and staggered federal approach to requiring electronic filing.

To set in motion the process of extracting the 990 data into usable form from the image files, both Public Resource (https://bulk.resource.org/irs.gov/eo/readme.html) and The Internet Archive (https://archive.org/details/IRS990) have obtained and posted openly, freely available DVDs of 990 data, obviating the need for others to request and buy them from the IRS. This should enable a wider ecosystem of those with the interest and resources to do optical character recognition to examine the data and work with it. Second, rather than relying on a single entity to step up and do the extraction, those funders who regularly spend millions each year on data extraction with several entities have the greatest incentive to organize the creation of a public repository of 990 data.

The data will be more reliable and trustworthy, and will be available more rapidly, if multiple entities are involved and the work is shared. Funders should consider convening a voluntary consortium of third parties, including those they already fund and other companies with the resources to assist with this project, as well as academics, who practice and study optical character recognition and other relevant technologies. Some parties will focus on extraction while others concentrate on redacting personally identifiable information and yet others will clean data. The Statistics of Income Division of the IRS (SOI) should be invited to participate and lend its expertise to the process. The SOI Division may find expanded public sector interest to be a reason to shift additional resources into the 990 program, which could allow expanded or more user friendly data releases and more in-depth analyses.

It isn't necessary to wait to have all the data in hand in raw form to start creating a bulk repository. The work can be done in phases to extract current tax returns and then work backwards to make the historic corpus available. In addition, certain parties extract particular fields of data already. For example, the National Center for Charitable Statistics compiles and cleans approximately 60 variables from the IRS Return Transaction File, which becomes its Core File that could be contributed and associated to the remaining fields once those are extracted by a third party.

Translating Open Data into Innovation

Thinking about how a scientist who never publishes reaps nothing, the sociologist Robert Merton wrote about the paradoxical way that information discovered by researchers only becomes theirs when they give it away. [84] So it is with data more generally — at least data, like the data culled from redacted 990's, that by law cannot be considered private or proprietary.

Below is a plausible sketch of how to begin maximizing the value of a nonprofit sector database platform. Such an entrepreneurial strategy could be adopted either by a nonprofit or a for-profit entity.

1. Secure tax or philanthropic funding for an institution that will compile and manage a nonprofit sector database with computable open data for the corpus of Forms 990 data. Identify a technical strategy for releasing the data.

2. Cultivate some marquee nonprofits and database users a la Glasspockets to help convince others to participate, too.

3. Invite all nonprofits to become shareholders or members by agreeing both to send the institution an electronic copy of the Form 990 they sent to the IRS and to check the accuracy of these data when it is posted.

4. List all nonprofits in the database regardless of whether they want to become shareholders or not by extracting information from their Form 990s.

5. Invite interested nonprofits to become premium members by paying to supplement their listings with keywords or other standardized information about themselves, their programs, or their impacts. Offer scholarships to those who want premium privileges but choose to provide services to the company like data checking rather than paying.

6. Make the database available under an agreement similar to a Creative Commons Attribution, Noncommercial, and ShareAlike license (by-nc-sa). This means: that licensees must credit the source; that they can copy, distribute, and modify the source for noncommercial purposes only; and that they can distribute it further only under the same licensing that governs the original. [85]

7. Negotiate agreements with commercial users of Form 990 data to purchase access to the database. If they pay a price no higher than what they spend now to compile their data, everyone wins. In particular, those commercial vendors cannot only continue selling their value-added services and functionality, but also expand them as part of a growing ecosystem. Or it may be possible to experiment with making the data free to everyone for any purpose, including commercial ones. With such a policy, vendors will have to earn profits on the functionality and services they provide rather than on the content.

8. Build and nurture an ecosystem of contributors and users who recognize their interdependent interests in seeing such a platform thrive.

For an example of an institution that has adopted a more entrepreneurial approach to providing nonprofit data, we need only look to CharityFocus.ca, a website opened in 2012 by Imagine Canada, a private organization "whose cause is Canada's charities and nonprofits." [86] The Canadian analogue of the Form 990 is the T3010. Although the Canada Revenue Agency (CRA) does not promote electronic filing of such forms, it does make the data from the T3010 (minus certain personal and protected information) available on its own website. In addition to providing search capabilities, the CRA also makes the entire database available for bulk downloading in Excel or other standard formats. The CRA joined with the RBC Foundation and Cenovus Energy to provide grants for the creation of CharityFocus by Imagine Canada.

The new website not only makes ten years of T3010 data easier to find, compare, and visualize, it also allows charities to upload more timely and detailed information about their programs and accomplishments. Public users have access to convenient links and tools, including the ability to make donations on the spot. Charities also have access both to regulatory information and to a "Quick Prep" service that helps them more accurately and completely fill out their T3010 forms.

Thousands of unique visitors have been recorded in each of the months since CharityFocus launched in February of 2012. There is no charge for using its data or services. Benefiting from a series of efforts like this to provide information and support, Canada's sector has apparently been thriving. According to the latest reports, total revenues of Canadian registered charities jumped from $192 to $207 billion between 2009 and 2010. Assets rose from $240 billion to $266 billion. About 10,000 of the 86,000 charities account for a very large portion of both revenue and assets.

LESSONS ABOUT OPENING DATA

- Innovate While Advocating: E-filing legislation is important but may be fraught with compromise. Traditional stakeholders are not necessarily fans of open data because it upsets the status quo and therefore cannot be counted upon as allies in the fight. In parallel to advocating for legislation, pilot projects with some data can demonstrate what's possible, help to win adherents, and justify the burden of legislation.

- Build a Wider Ecosystem: Incumbent stakeholders and institutions, while often most knowledgeable about a sector, are not always those who understand or support openness. Instead, open data advocates must reach outside the sector to those who are knowledgeable about open data in tech and media to uncover benefits and develop innovations.

- Data in Any Form are Better Than No Data: Even image files like the 990s should be uploaded as is in a public, online repository and made freely and publicly available. There will always be someone with the wherewithal to make the data more useable and derive insight from it. Don't wait!

- Open Data Isn't Enough: Data liberation has to serve the goal of solving problems and improving the human condition. Hence it isn't enough to think about how to make data open. Rather, work and thinking must be expended to ensure that raw data translates into impact. That means it is as important to make sure innovative people know about the availability of data as they are to make it available.

Figure 10 – Lessons about Opening Data

Alone, open government data — even open 990 data — produce less value than an "innovation ecosystem" where people inside and outside of government with a variety of skills examine, poke, hack, map, visualize, and mash up the data to gain insights and solve problems. Once the data are liberated from image files into standard, open, computable formats, then the "real" work begins. Given the millions that can be saved on data extraction if the data are either produced electronically (in the long run) or extracted once for shared use (in the short run), effort and funds should be available to spur and invest in innovations. This section sets out a recommended action plan for creating an "innovation ecosystem" together with further estimates of costs and risks.

GET A DATABASE OUT THERE

It is important to put a substantive, computable Forms 990 database online sooner rather than later in order to generate interest in the 990 data and spur innovation.

There are various options for doing this. The process need not be sustainable as long as it starts the ball rolling on the development of innovations, including interesting apps and services. In fact, it need not even be comprehensive for now. A computable, open, and high-quality dataset covering, say, the largest 25,000 nonprofits could be enough to kickstart both innovative applications and the demand for more such data. A thematic slice of a subsector, such as a dataset describing all the nonprofits working on a given topic like the environment, homelessness, or health issues, could have a similar effect.

As described above, the IRS is researching the possible release of computable data from its Master File on an experimental basis in 2013. The research is investigating the release of many but not all fields for a large sample of nonprofits, including all of those above certain thresholds. This is a positive development, and one that could prove very significant. If the community puts this to good use, such a dataset might become a significant and continuing resource both because of the information it contains and because of the applications it could inspire.

HOLD DATA DIVES AND CODE-A-THONS

A relatively quick way to test the data and to begin generating results and buzz is to hold a one-off event. Unlike competitions, which encourage the development of tools over an extended period, Code-a-Thons (also called Hackathons or Mash-a-thons) are short events that range from a few hours to a few days or even months. They bring together developers, entrepreneurs, academics, and experts in the field in one place to devise and execute on the creation of new tools.

A Code-a-Thon differs from a more traditional conference in its deliverables. Extended discussion is eschewed in favor of giving technologists and coders the material to develop solutions on the spot. Because of the short time frame, these solutions are often incomplete and show only a portion of what the overall application or tool can become. Organizations sometimes continue working with Code-a-Thon participants to continue developing promising tools after the event or work on their own to develop new businesses or simply tools that they give away or sell in online App stores. In the Code For Change [87] model of Hackathon run by New York University's Robert F. Wagner Graduate School of Public Service, government institutions and nonprofits articulated problems they needed to solve and worked together with teams of coders over an extended period to develop innovative tools informed by institutional expertise.

A LOOK INTO A HACKATHON ECOSYSTEM: HEALTH 2.0

The Health 2.0 ecosystem is the blueprint that other open data movements base their efforts on. The ecosystem is built around a robust set of data and an active community outreach program that engages a number of satellite hackathons. This community has scaled up and grown from a number of interested parties to an evolving community that promotes problem solving by cultivating relationships.

The movement began with ideation around health data, which brought together data providers and data consumers. This group looked at the data formats, what information was included in the data, and what could be done with the data. Their work led to a number of hackathons where apps and research tools are created around a particular dataset or presented problem. These apps and tools are refined over time and the outreach programs build towards the annual Health Datapalooza, the flagship program of the Health 2.0 movement. The event gathers a large group of stakeholders and innovators to evangelize and celebrate the community. Apps are demonstrated at the event and presentations are given on "Action Beats," which generally involve organizational or challenge announcements.

The success of the Health 2.0 movement has relied on the partnerships that have been cultivated over months and years, but also on the lives it has saved. With an increase in transparency, statistics on the number of instances where hospital staff provided antibiotics to patients requiring surgery increased as that data became open. This level of transparency, in no uncertain terms, saved lives.

Figure 11 – Health 2.0

Code-a-Thons are quick ways to make demonstrable progress on developing tools and solutions. Even more important, they are a way to convene and forge community around shared interest. The opening up of data about nonprofits will pique the interest of organizations and people not previously thought of as "stakeholders." We spoke with a group of tech savvy civic activists from Code for America who showed tremendous interest in nonprofit data and were eager to help cultivate an ecosystem around these data. By marrying technology-backed expertise with the knowledge of existing stakeholders, Code-a-Thons help to bring new skill sets and abilities together to accomplish good work.

Code-a-Thons produce innovative ideas, but like an app contest, the products will often be disconnected and unresponsive to the greatest needs (except in the Wagner model). They're also limited to what can be accomplished in one or two days. Products from Code-a-Thons may need to be further developed, requiring additional funding and the willingness of volunteers to continue working in their spare time. In addition, unlike stakeholders who hire professional vendors, volunteer coders may have only a limited knowledge of the field.

GUIDESTAR AND THE DATA DIVE

Chuck McLean, Vice Present of Research at GuideStar USA, spoke about his experience at the DataKind "Data Dive," a two-day Code-a-Thon to create data-backed tools for nonprofit financial analysis. The event was co-hosted with Independent Sector, the coalition of foundations and nonprofits. Prior to the Data Dive, Mr. McLean's expectations were not high going into the event. The data would be complicated. People wouldn't know what to do with it. There wouldn't be enough time to get anything done.

The Data Dive consisted primarily of a coding session where technologists worked on software in response to challenges posed by presenters, including Mr. McLean. The technologists worked from two data sets provided by GuideStar, including Forms 990 data. Mr. McLean presented the group of technologists with a straightfor-ward request for a new tool. "What we were basically looking for is an early predictor of financial failure on the part of a nonprofit," said Mr. McLean. At the end of the second day, coders presented their work to the group and judges.

One group created a promising, regression-based prototype and offered to continue working on the tool with GuideStar on a voluntary basis. When asked how his view of events like the Data Dive changed, Mr. McLean said he came out of the experience thinking it was a great idea. "I've never had an experience with crowd-sourc-ing or open source where instead of just one person doing their proprietary thing, lots of different people are looking at it and tweaking it," he said. "I am definitely a convert to that way of doing things."

Figure 12 – GuideStar and the Data Dive

ORGANIZE COMPETITIONS AND PRIZES [88]

Supported by recent legislation [89] and Obama Administration policy, [90] prize-backed competitions involv-ing public sector data have become increasingly commonplace. [91] In 2008, the Office of the Washington, DC, Chief Information Officer sponsored the first Apps for Democracy competition followed in 2009 by the Sunlight Foundation's Apps for America competition [92] to create useful tools using open government data. As in these cases, sometimes the competitions are theme-based (i.e., develop an app that improves transparency) without specifying the data to be used. New York City has held three "Big Apps" contests, which encourage use of data made avail-able by the city government in machine-readable formats.

In other cases, the competition revolves around a common dataset. The Community Health Data Initiative of the Department of Health and Human Services has sponsored "datapaloozas," events designed to encourage the develop-ment of tools to improve wellness using their datasets specifically. The Chief Technology Officer of the U.S., Todd Park, has now sponsored datapaloozas in energy and other sectors as well. First Lady Michelle Obama lent her name and support to a contest to create the best healthy eating games using the Food Pyramid and the associated United States Department of Agriculture dataset on the nutritional values of the 1000 most commonly consumed foods. [93]

Competitions provide a relatively low-cost way to spur innovation. [94] They generate enthusiasm and attract par-ticipation from those outside of government. Occasionally, there are celebrity judges as in the case of the Jon Bon Jovi app for veterans competition. [95] The literature on two-sided markets points to the importance of marquee contributors, users, and success stories in launching a platform.

Competitions often result in unanticipated innovations, such as the creation of an "Asthmopolis" tool to map and track incidences of asthma inhaler use as a way to pinpoint where asthma flares occur, or the creation of a new, more readable and accessible version of the Federal Register. In both cases, these tools responded to long-felt but unarticulated needs.

However, poorly designed app competitions produce a hodgepodge of unrelated or repetitive apps, such as the many "When is my bus coming?" apps inspired by the release of transportation data by municipal transit authori-ties. Without direction, app competitions may produce results that do not respond to the highest priority needs. When contests deal with easy to understand and well-known topics like transit data, they often attract widespread participation and varied submissions. On more esoteric topics, it takes effort to alert experts to the opportunity to

participate and to explain to those unfamiliar with the field the kinds of problems that need solving. While not all contests are app contests, most open government data events do appeal to data-savvy technologists. Many in the nonprofit community who understand the challenges well are not data-literate. Collaboration is a must for creating a rich ecosystem of innovation.

In addition, while app contests usually produce working prototypes, these are not the same as completed products. Further development is required and it is in the best interests of the organizers to set up a method to continue working with the developers after the competition is completed.

CREATE AN APP STORE

By themselves, hackathons and data dives are often one-off events that can lead to the development of half-baked and duplicative iPhone apps, rather than sustained innovation in the sector. In parallel to creating events to generate buzz and excitement and the development of new tools, there need to be strategies for refining the tools in conversation with experts in the sector as well as distribution strategies to help with marketing whatever gets developed. Hence there's a need for an "app store" for nonprofit innovations — a central showcase that curates the "best of" nonprofit tools. An app store might also provide a way to sustain innovation in the sector by creating a source of revenue for innovators by selling their apps. One can imagine one or more of the larger players in the nonprofit space, such as GuideStar, the Urban Institute, or the Foundation Center, hosting such an app store.

RE-INVESTIGATE LONGER-TERM SOLUTIONS

As many innovators learn, people do not always know what they want until they see it. Once demonstrations have begun firing imaginations, it makes sense to reassess the alternatives for a long-term solution. Formal and informal surveys would then be more informative about the support for different kinds of 990 database platforms — whether existing or new, for-profit or not. The odds of a legislative win might change as well with greater publicity.

BUILD COMMUNITY AND NURTURE AN ECOSYSTEM

For the 990 data to generate value to the wider public they need these to be available in computable form for free. The primary rationale for liberating data is to get more people working with and looking at the data. We want people to be able to use the corpus of the data in its entirety to facilitate cross-industry analysis rather than single entity lookup. We want more people developing innovations using the data, such as tools to identify where a homeless vet can get services in a given state. Finally, we believe that without free bulk downloads, we are restricted to an ecosystem of well-resourced players who can afford to buy the data to build the tools that interest them or that they can sell to a large enough audience. The goal, instead, should be to reach the much wider potential audience of innovators who can use this public resource for the greater public good. Put another way, if we don't give access to the data to smaller groups and individuals who could use the data for good, including using them to assess the quality of the data, we miss the "long tail" of opportunity for innovation.

Complicating the story is that unlike some datasets, such as the Federal Register, liberating these data isn't a matter of finding the political will and flipping a switch. The 990 data need work to make the data better organized, more timely, and more accurate. The most interesting 990 data are neither structured nor standardized: They are contained in the attachments, which can come in any format. Attachments can be completely free form lists of grants, for example, in a Word document rather than a spreadsheet.

There is a real start-up cost to creating open 990 data. It is a natural assumption to want to recover those costs from the end users. But, in fact, we believe that the end users are going to be the real source of energy for cleaning, correcting, and innovating with the data. Hence the costs need to be borne elsewhere. Right now, that means we need a coalition of funders and groups to extract the data and innovators interested in working with the data to demonstrate their value. This solution may not be sustainable forever. Rather, high impact pilots like this, if they are successful, will help demonstrate the value of open 990 data whether or not the original funders continue to support it.

Acknowledgments

Thank you to the many people who gave of their time and expertise for this paper. These include, but are not limited to, the list of people below. For additional interviewees see appendix 7.

Telephone interview with Sophie Raseman, Director of Smart Disclosure and Senior Policy Adviser at the U.S. Department of Treasury (February 29, 2012).

In person interview with Bill Nichols, U.S. Department of Treasury (March 14, 2012).

Telephone interview with Lois Lerner, Director of Exempt Organizations Division and, David Fish, Group Manager at the Exempt Organization Tech Division at the Internal Revenue Service (March 23, 2012).

Telephone interview with Darin Mckeever, Deputy Director at the Bill & Melinda Gates Foundation (March 24, 2012)

Telephone interview with Ruth Madrigal, Attorney Adviser at the Office of Tax Policy in the U.S. Department of Treasury (March 29, 2012).

In Person interview with Hudson Hollister, Executive Director at Data Transparency Coalition (March 31, 2012).

Telephone interview with Katie Roeger, Research Associate and Assistant Director at the National Center for Charitable Statistics at the Urban Institute (April 12, 2012).

Telephone interview with Cindy Lott, Senior Counsel to the National State Attorneys General Program at Columbia Law School (April 16, 2012).

Telephone interview with Jennifer Deger, Controller at the Bill & Melinda Gates Foundation (April 17, 2012).

Telephone interview with Tom Pollak, Senior Research Associate and Program Director for the National Center for Charitable Statistics at the Urban Institute (April 19, 2012).

Telephone interview with Elizabeth Boris, Director at the Center on Nonprofits and Philanthropy at the Urban Institute (April 27, 2012).

Telephone interview with Sean Delaney, Executive Director of the Lawyers Alliance for New York (April 27, 2012).

Telephone interview with Jacob Harold, former Program Officer at the William and Flora Hewlett Foundation (April 29, 2012).

Telephone interview with Chuck McLean, Vice President of Research at GuideStar (May 07, 2012).

Telephone interview with Emmy So, Database Operations Assistant at the Foundation Center (May 10, 2012).

In Person interview with Joel Gurin, former Chair of White House Task Force on Smart Disclosure (May 11, 2012).

Telephone interview with Lucy Bernholz, Visiting Scholar at Stanford University Center on Philanthropy and Civil Society (May 11, 2012).

Telephone interview with Jon Dunford at the Urban Institute (May 11, 2012).

Telephone interview with Jonathan Greenblatt, Director of the White House Office of Social Innovation and Civic Participation (May 11, 2012).

Telephone interview with Mike Wash, Chief Information Officer at the National Archives (May 18, 2012).

In Person interview with Bradford Smith, President of the Foundation Center (May 29, 2012).

Telephone interview with Hugh Jones, Supervising Deputy Attorney General Tax and Charities Division at the Hawaii Department of the Attorney General (May 29, 2012).

Telephone interview with Christine Aube, Director of Data Management at GuideStar (May 29, 2012).

Telephone interview with Matt Hampel, Ben Sheldon, and Serena Wales, Code for America Fellows (May 30, 2012).

Telephone interview with Bob Ottenhoff, former President and CEO of GuideStar (June 06, 2012).

Telephone interview with Stacy Palmer, Editor, *Chronicle of Philanthropy*, (June 7, 2012).

Telephone interview with David Fish, Group Manager of Exempt Organizations at the IRS (June 15, 2012).

Telephone interview with Bob Carlson, Assistant Attorney General at the Missouri Attorney General's Office (June 27, 2012).

Telephone interview with Lin Smith, Managing Director of the National Economics and Statistics Tax Sampling Group at PricewaterhouseCoopers (June 28, 2012).

Telephone interview with Karen Gano, Assistant Attorney General at the Connecticut Office of the Attorney General (June 29, 2012).

Telephone interview with Elizabeth Korsmo, Assistant Attorney General at the Office of New Mexico Attorney General (July 03, 2012).

Telephone interview with Susan Boehmer, Director, and Barry Johnson, Chief Special Studies Branch, of the Statistics of Income division at the IRS (October 22, 2012)

ENDNOTES

1. Quick Facts About Nonprofits, http://nccs.urban.org/statistics/quickfacts.cfm (last visited Aug. 2, 2012).
2. The family of Forms 990 includes the Form 990, the Form 990-PF, the Form 990-EZ, and the Form 990-N. The forms each have specific thresholds for the kinds of nonprofit organizations that are required to file. Nearly all nonprofits are required to file at least one type of form. (See Appendix 2: The Federal Forms for more details.)
3. Statement made by Diana Aviv, CEO of the Independent Sector, to the House Ways and Means Committee (May 16, 2012).
4. Publicity of information required from certain exempt organizations and certain trusts, 26 U.S.C. § 6104 (1999).
5. Beth Simone Noveck, Open Data: The Democratic Imperative, in *Crooked Timber* (July 5, 2012), at http://crookedtimber.org/2012/07/05/open-data-the-democratic-imperative/.
6. Matt Parker, Lamplight Database Systems, in Kate Hodge, Best Bits: Open Data for Charities, *the Guardian* (May 25, 2012), http://www.guardian.co.uk/voluntary-sector-network/2012/may/25/open-data-charities-advice.
7. See, e.g., Nonprofit Almanac 2012 at http://www.urban.org/books/nonprofit-almanac-2012/index.cfm.
8. See for example http://www.irs.gov/uac/SOI-Tax-Stats-Charities-and-Other-Tax-Exempt-Organizations-Statistics or http://nccs.urban.org/database/overview.cfm.
9. Bradford Smith, Philanthropy Data's Dilemma, July 9, 2012, at http://pndblog.typepad.com/pndblog/2012/07/philanthropys-data-dilemma.html.
10. Noelle Barton, IRS Warns Charities to Avoid Disclosing Social Security Numbers, *Chronicle of Philanthropy*, April 25, 2012, available at http://philanthropy.com/article/IRS-Warns-Charities-to-Avoid/131655/. ("'Because the law doesn't allow the IRS to redact Social Security numbers when we make 990s public, it's important for organizations, preparers, and you to make sure this kind of information isn't put on the form, which creates a risk of identity theft,' said Lois Lerner, director of the IRS division that oversees charities, in a speech last week at Georgetown University Law School.")
11. In June 2011, the IRS announced that approximately 275,000 organizations had their exempt-status revoked, http://www.irs.gov/uac/IRS-Identifies-Organizations-that-Have-Lost-Tax-Exempt-Status%3B-Announces-Special-Steps-to-Help-Revoked-Organizations
12. See www.greatschools.org and www.dataqualitycampaign.org.
13. Digital Government: Building a 21st Century Platform to Better Serve the American People (May 23, 2012), available at www.whitehouse.gov/digitalgov/pdf.
14. IRM Section 4.75.10.4-1(A).
15. IRM Section 4.75.10.4-1(B).
16. IRM Section 4.75.10.4-1(C).
17. IRM Section 4.75.10.4-1(D).
18. http://www.irs.gov/charities/article/0,,id=139480,00.html; Form 1023, Part 11.
19. Capitol Hill Hearing Testimony, House Ways and Means, April 20, 2005; http://www.gao.gov/assets/120/111547.html.
20. For the history of the 2008 redesign, see Chronological History: Redesign of the 2008 Form 990 and Corresponding Instructions, available at http://www.irs.gov/Charities-&-Non-Profits/Charitable-Organizations/Chronological-History:--Redesign-of-the-2008-Form-990-and-Corresponding-Instructions.
21. William H. Byrnes, The Private Foundation's Topsy Turvy Road in the American Political Process, 4 Hous. Bus. & Tax L.J. 496 (2004).
22. Revenue Act of 1943 Section 117(a), 26 U.S.C. Section 54 (1944).
23. Paul Arnsberger, Mellissa Ludlum, Margaret Riley, and Mark Stanton, A History of the Tax-Exempt Sector: An SOI Perspective.
24. Diane L. Fahey, Taxing Nonprofits Out of Business, 62 Wash. & Lee L. Rev. 547 (2005).
25. Paul Arnsberger, Mellissa Ludlum, Margaret Riley, and Mark Stanton, A History of the Tax-Exempt Sector: An SOI Perspective.
26. 1965 Treasury Report.
27. Thomas A. Troyer, The 1969 Private Foundation Law: Historical Perspective on Its Origins and Underpinnings, the Exempt Organization Tax Review Jan. 2000 Vol. 27 No. 1.
28. Tax Advisory Group of the Office of Legal and Regulatory Affairs, American Health Lawyers Association, The Public Charity Status of Hospitals and Health Care organizations Under the "Operated in Connection With" Relationship Test of Section 509(a)(3) of the Internal Revenue Code of 1986, 22 AHA. Jour. Of Health Law (1989).
29. IRC Section 6033(i).
30. George K. Yin, An Overview of the Tax-exempt Sector, 109 CIS Legis. Hist. P.L. 280 (109 PL 280) (2005).
31. U.S.C. Section 1223(a) (2006).
32. IRC Section 6033(j).
33. Frederick J. Gerhart, Charitable Incentives and Charitable Reforms Under the Pension Protection Act of 2006, 19 Health Law. 17 3 (2007).
34. Evelyn A. Lewis, Charitable Waste: Consideration of a "Waste Not, Want Not" Tax, 30 Va. Tax Rev. 39 (2010).
35. The hearing was held on July 25, 2012 and focused on "organizational and compliance issues related to public charities, including the increased complexity of public charity organizational structures, the rules governing profit-generating activities giving rise to unrelated business income tax, and whether the newly redesigned Form 990 is promoting increased compliance and transparency." See http://waysandmeans.house.gov/news/documentsingle.aspx?DocumentID=303615.
36. Karen Donnelly, Good Governance: Has the IRS Usurped the Business Judgment of Tax-Exempt Organizations in the Name of Transparency and Accountability?, 79 UMKC L. Rev. 163 (2010).
37. Summary of the Form 990 Redesign Process.
38. Ibid.
39. Memorandum on Transparency and Open Government, January 21, 2009, available at http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment.
40. Data.Gov, http://www.data.gov/communities (last visited Aug. 2, 2012).
41. Unlike Commonwealth countries that assert Crown Copyright in government works, United States copyright law provides that the federal government cannot assert copyright (17 USC §105). The absence of a national copyright has accelerated the move to open data

in America. In countries like the UK, the national government retroactively applies a license known as the Open Government License to make works technically owned by the Queen, such as legal codes and government information, freely usable to all.

42. Robinson, David G., Yu, Harlan, Zeller, William P. and Felten, Edward W. Government Data and the Invisible Hand. 2009. *Yale Journal* of *Law & Technology*, Vol. 11, p. 160.

43. Enabling third parties to download data from government servers needs to be done in such a way as to minimize risk to critical infrastructure.

44. Tax Information for Churches and Religious Organizations - Filing Requirements, http://www.irs.gov/charities/churches/article/0,,id=132081,00.html (last visited Aug. 2, 2012).

45. Internal Revenue Service, Which Forms Do Exempt Organizations File, http://www.irs.gov/charities/article/0,,id=184445,00.html (last visited Aug. 2, 2012).

46. Internal Revenue Service, Instructions for Form 990, http://www.irs.gov/instructions/i990/ar01.html (last visited Aug. 2, 2012).

47. IRS Releases Discussion Draft of Redesigned Form 990 for Tax-Exempt Organizations, IRS.gov, http://www.irs.gov/newsroom/article/0,,id=171329,00.html (last visited).

48. http://www.irs.gov/efile/article/0,,id=108211,00.html; IRC Section 6033(i).

49. Peter Swords, Nonprofit Coordinating Committee of New York, "How to Read the IRS Form 990 & Find Out What it Means," available at http://www.npccny.org/Form_990/990.htm.

50. The Secretary of the Treasury may not require "any person" to file their return electronically unless they file at least 250 returns during the calendar year and the must take into account the taxpayers ability to comply, at reasonable cost, with the regulation. IRC Section 6011(e)(2)(A); IRC Section 6011(e)(2)(B).

51. Internal Revenue Service, Internal Revenue Manual, Part 3, ch. 3.20.12.3.7.1 (01-01-2012) (available at http://www.irs.gov/irm/part3/irm_03-020-012r.html#d0e1731).

52. This interview was conducted on June 4, 2012. See list of interviews at Appendix 8 Internal Revenue Service, http://www.irs.gov/instructions/i4506a/ch01.html (last visited Aug. 2, 2012).

53. TIFF, which stands for Tagged Image File Format, is a file format for storing images. http://en.wikipedia.org/wiki/Tagged_Image_File_Format.

54. IRS, Avoiding Identity Theft — Personal Identity Information on Exempt Organization Returns, http://www.irs.gov/charities/article/0,,id=258203,00.html (last visited on Aug. 2, 2012).

55. The machine readable data are at the bottom of each page under "Microdata files." The first is for charitable organizations and the second is for private foundations. See http://www.irs.gov/uac/SOI-Tax-Stats-Charities-&-Other-Tax-Exempt-Organizations-Statistics and http://www.irs.gov/uac/SOI-Tax-Stats-Domestic-Private-Foundation-and-Charitable-Trust-Statistics.

56. http://usaspending.gov/explore?tab=By+Prime+Awardee&typeofview=complete&pageno=2&fiscal_year=all&piid=TIRNO09C00070

57. Interview with Emily So, the Foundation Center, on May 10, 2012.

58. The GuideStar contract is listed on USASpending.gov.. http://usaspending.gov/search?form_fields={%22search_term%22%3A%22INTERNAL+REVENUE+SERVICE%22%2C%22recipient_name%22%3A%22GuideStar%22}.

59. Memorandum released by the Foundation Center, foundationcenter.org/findfunders/IRS_990_notice.pdf. Estimates derived from calculations made by Katherine Toran, Research Assistant at the Urban Institute.

60. The SOI Bulletin also includes separate annual articles for organizations exempt under Internal Revenue Code sections 501(c)(3)-(9) and for private foundations. The articles, published by tax year, describe the financial state of these organizations. The analysis classifies tax-exempt organizations by asset size and subsection code.

61. Estimates derived from calculations made by Katherine Toran, Research Assistant at the Urban Institute.

62. Givewell, http://givewell.org/ (last visited Aug. 2, 2012).

63. For most purposes, it is best to analyze data by tax year, designated by the year in which an organization's accounting period begins. Organizations are allowed to select any calendar month as the start of their accounting period, therefore accounting periods for a given tax year can vary by up to 12 months. Most Form 990-series returns are filed within a 12-month period after the close of each organization's accounting period. Automatic and other extensions can extend the filing period by up to 6 additional months. As a result, a sufficiently complete population of returns cannot be expected until a full two years after the tax year. Processing of the return information of the later-filed returns requires additional time. That schedule suggests the minimum lag between the end of a tax year and the possible availability of comprehensive return data for that year in analytical form is closer to 30 months.

64. Johns Hopkins Center for Civil Society http://ccss.jhu.edu/.

65. Hugh Jones Interview, May 29, 2012.

66. Code for America Interview, May 30, 2012.

67. http://articles.latimes.com/2012/oct/01/nation/la-na-nn-harvard-attorney-scam-navy-vets-20121001.

68. http://storify.com/mccabe_dan/can-we-help-GuideStar-predict-nonprofit-success-or.

69. GuideStar is just one organization that offers a Nonprofit Compensation Report for various prices, primarily dependent on the number of users who will have access to the information. http://www.GuideStar.org/rxg/products/nonprofit-compensation-solutions/GuideStar-nonprofit-compensation-report.aspx.

70. Johns Hopkins University - The Johns Hopkins Economic Data Project, http://ccss.jhu.edu/research-projects/nonprofit-economic-data (last visited Aug. 2, 2012) .

71. Nate Herpich, Google Co-founder Sergey Brin Partners with The Michael J. Fox Foundation to Create a 'Manhattan Project' in Search for Parkinson's Cure, *Bloomberg Businessweek* (May 11, 2012), http://blog.michaeljfox.org/2012/05/bloomberg-businessweek-google-co-founder-sergey-brin-partners-michael-j-fox-foundation-create-manhattan-project-search-parkinsons-cure/.

72. Molly Peterson, IRS Struggles to Know if Big Nonprofits Pay the Taxes They Owe, *Chronicle of Philanthropy* (July 25, 2012) http://philanthropy.com/blogs/government-and-politics/irs-struggles-to-know-if-big-nonprofits-pay-the-taxes-they-owe/30575?cid=pw&utm_source=pw&utm_medium=en.

73. Citizens United v. Federal Election Commission, 558 U.S. 310 (2010).

74. We spoke with Sheila Krumholz on June 5th, 2012.

75. http://en.wikipedia.org/wiki/Legal_Entity_Identification_for_Financial_Contracts.

76. Public Resource submitted a grant application to the Knight Data Challenge in 2012, requesting money to "put 10 years of IRS Form 990, 990-PF, and 990T online in bulk, extract 75 million fields of data." http://newschallenge2.tumblr.com/post/25545821455/opening-irs-nonprofit-tax-returns.

77. http://efile.form990.org/.

78. http://www.irs.gov/Charities-&-Non-Profits/Copies-of-Scanned-EO-Returns-Available.

79. PublicResource.org acts as a kind of "Robin Hood" of government data except they don't steal. They raise the money to buy the data at full price and then post it on line for anyone to download for free. In the case of the 990 data, they bought and received Forms 990 CD/DVDs from the IRS via the White House via the GSA. Public Resource took the TIFF files and organized them as PDFs with associated metadata. The organization has made the family of Forms 990 available as PDFs from 2002 to 2012.

80. Currently, PDFs of Forms 990, 990-EZ, 990-PF, and 990-T are available from 2002 to September, 2012. More information can be found here: https://bulk.resource.org/irs.gov/eo/readme.html.

81. See also http://boingboing.net/2012/11/01/tax-returns-for-6461326-tax.html.

82. The Modernized e-File system allows nonprofits to file their Forms 990 online in an electronic, structured format. The system was rolled out in 2004 and since its inception has allowed for the voluntary and mandatory filing of Forms 990 online. See http://www.irs.gov/uac/IRS-E-File:-A-History.

83. The Glasspockets initiative makes grant award data available for 15 of the largest foundations in America. These organizations post their data online and the Foundation Center collects these data using an application programming interface (API). This API combines the underlying data and makes them queryable. More can be found at http://foundationcenter.org/media/news/20121009.html.

84. Merton, Robert K, The Unanticipated Consequences of Purposive Social Action, *American Sociological Review* 1 (6): 894–904

85. http://creativecommons.org/licenses/by-nc-sa/2.5/.

86. http://www.imaginecanada.ca/.

87. http://wagner.nyu.edu/labs/codeforchange/.

88. Open Government Directive section on prizes; Guidance on the Use of Challenges and Prizes to Promote Open Government, 03/08/2010, Jeffrey D. Zients, EOP OMB.

89. Challenge.gov, http://challenge.gov (last visited on August 2, 2012).

90. America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act of 2007, P.L. 110-69

91. William Neuman, Nutrition Plate Unveiled, Replacing Food Pyramid, the New York Times, June 2, 2011 http://www.nytimes.com/2011/06/03/business/03plate.html.

92. The Sunlight Foundation, http://sunlightlabs.com/contests/appsforamerica2/ (last visited on August 2, 2012)

93. Competitions held through Recipesforkidschallenge.com are promoted by Challenge.gov, a central platform run by the Federal government to engage citizens in solving larger governmental problems.

94. Nicole Wallace, Competitions Seek Examples of Nonprofit Innovation, *Chronicle of Philanthropy* (June 1, 2012) http://philanthropy.com/blogs/innovation/competitions-seek-examples-of-nonprofit-innovation/922?sid=pt&utm_source=pt&utm_medium=en.

95. Heidi Williams, Innovation Inducement Prizes: Connecting Research to Policy, *Journal of Policy Analysis and Management*, Vol. 00, No. 0, 1–25 (2012).

Information for Impact: Liberating Nonprofit Sector Data

APPENDICES

APPENDIX 1: The Nonprofit Sector

Highlights:

- The nonprofit community is a diverse, eclectic mix of over 1.5 million organizations that provide various services to constituents. [1]

- Nonprofits collect trillions of dollars in total revenue and hold trillions in assets. [2]

Analysis:

Nonprofit organizations are able to act as vital tools in filling community needs because of their inherent flexibility in advancing the public good. Nonprofit organizations vary in the constituents they serve and receive tax-exempt status for providing services under certain classifications. The Internal Revenue Code (the "IRC") establishes approximately 30 types of recognized nonprofit classifications. [3] These include entities that are specifically organized for "religious, charitable, scientific, testing for public safety, literacy, or educational purposes." [4] In 2009, the nonprofit community "accounted for 9% of all wages and salaries paid in the United States," and its share of the GDP was 5.4%. [5]

The largest number of organizations fall under IRC Section 501(c)(3). A 501(c)(3) organization is either considered a "Public Charity" or a "Private Foundation." Public Charities are "the generic term given to the excepted organizations." [6] Private Foundations are not defined, but rather are classified as organizations that do not fall under a set of four enumerated exceptions. [7]

Public Charities represent a large portion of the nonprofit community and are predominant asset holders and revenue receivers in the field. In 2009, Public Charities reported $1.40 trillion in total revenues and $2.6 trillion in assets. [8] This amount of total revenue and total assets is more than either private foundations or other nonprofits retain. [9]

A small number of organizations, particularly in the health and higher education fields, receive a disproportionately high share of all Public Charities' revenue. While hospitals represented less than 1% of all Public Charities, they collected 41% of total revenues reported in filings reviewed for a recent study. [10] Health charities held 36% of all Public Charity assets. Similarly, higher education institutions represented only 0.5% of Public Charities but accounted for more than 11% of total revenue. [11] Education charities made up 18% of charities, but held approximately 33% of total Public Charity assets. In contrast, nonprofits in the arts, culture, and humanities made up over 12% of the Public Charities, but only generated 2.3% of the revenue. [12]

Nonprofits collect revenues from multiple sources and through various means. The nonprofit community usually collects revenue in the form of payment for services, government grants, private contributions, or through returns earned on investments. In 2005, Public Charities raised a total of $1.2 trillion in revenue, of which: 49% came from payment of services, which includes medical care or tuition; 29% came from government grants or payments, which does not include subsidies; 12% came from private contributions, which include individual gifts, bequests, corporate donations, and foundation donations; 7% came from investment income; and 3% came from other sources of revenue, such as "membership dues, net special events income, and other miscellaneous revenue-raising activities." [13] The government contributes more in grants and payments for health and human services than for any other type of Public Charities. Regarding overall grants and subsidies provided to the charitable sector, the Congressional Research Service stated that in 2009:

Estimates regarding the value of the government's relationship with the charitable sector suggest that grants amount to about $100 billion, with the federal government supplying about 90% of the funds. Federal tax subsidies are valued to be approximately $115 billion to $130 billion, and state and local tax subsidies are approximately $30 billion to $50 billion. In sum, the government provides approximately $245 billion to $280 billion to the nonprofit and charitable sector via grants and tax subsidies. [14]

APPENDIX 2: The Federal Forms

Highlights:

- The Forms 990 are a family consisting of the Form 990, Form 990-PF, Form 990-EZ, and Form 990-N. Each has certain thresholds that establish which nonprofit types must file which form.

- The world of the Forms 990 has undergone two recent changes: a major redesign of the Form 990, and the requirement for small nonprofit organizations to file the electronic Form 990-N.

- The Forms 990 must be reviewed by the IRS, but do not create revenue for the Treasury. Any momentum for change within the agency must contend with the fact that innovations may increase workloads in the short-term without providing new sources of revenue.

Analysis:

Information on nonprofits comes from multiple information streams, but much of it supplements the information within the Forms 990. The Internal Revenue Service acts as the tax administrator for the federal government and has regulatory power over the nonprofit community. The IRS has two primary responsibilities in overseeing nonprofits: reviewing applications for the dispensation of tax-exempt status and examining annually filed Forms 990. The Forms 990 require that a nonprofit disclose information about the organization and its transactions, including the nonprofit's major service programs, its officer salaries, and its balance sheet. According to the Internal Revenue Manual (the "IRM"), the primary objectives for examination of annually filed Forms 990 is to determine if the organization is operating in a manner consistent with its exempt purpose, [15] if the forms are complete, [16] if the organization filed all forms and information required of it, [17] and if there are any outstanding taxes the organization must pay (nonprofits sometimes have associated taxable, for-profit activities). [18] Violations may result in enforcement in the form of excise taxes or a revocation of tax-exempt status. However, the information provided within the Forms 990 exceeds that which the IRS needs to conduct this task and represents a valuable, publicly available source of information on the nonprofit community.

The Forms 990 consist of the Form 990, Form 990 EZ, Form 990-PF, and Form 990-N. The rules for determining what type of organization must file each form, normally depend both on how much the organization earns in revenue and on organization type. Each form varies in length and information requested.

The Form 990 is an annual filing form that may be filed by virtually any nonprofit that is not a private foundation. This includes certain political organizations, social welfare and charity organizations, and charitable trusts. The form is broken up into 12 different parts spread over 12 pages. It provides over 1,000 fields that may be checked or require a numerical value input or specific text-based information. These disclosures include, but are not limited to, basic directory information, [19] a calculation of revenue and expenses, [20] information on the organizational structure, [21] a description of the organization's mission, any new services being offered, any losses of services formerly offered, and a description of the three largest services offered by the organization. [22] Organizations with incomes equal to or greater than $200,000 or assets that are equal to or greater than $500,000 must file the Form 990. However, other organizations may file this form should they choose to do so. In 2010, there were 261,016 Form 990 filers, 76,996 of which filed electronically. [23]

The Form 990-PF is similar to the Form 990, but may only be filed by private foundations. The Form 990-PF requires disclosure of, among other things, basic directory information, [24] an analysis of revenue and expenses, [25] a description of assets, liabilities, and net assets, [26] and capital gains or losses on investment income. [27] This 13-page form is broken up into 17 different parts and consists of over 900 fields that may be checked or require a numerical value input or specific text-based information. In 2010, there were 112,939 Form 990-PFs filed, 12,524 of which were filed electronically. [28] One reason for the lower percentage of private foundations filing electronically is that much of their required information is submitted in attachments, which do not lend themselves easily to electronic filing.

The Form 990-EZ is a convenience created to ease the burden for medium-sized nonprofits. What distinguishes this form from the Form 990 is the requirement that the filing organization earns less than $200,000 in gross receipts and has total assets less than $500,000 at the end of the filing year. The 990-EZ is a shorter form, four pages in length, and only has 6 parts and over 300 fields that may be checked or require numerical value input or specific text-based information. It requests, among other things, basic directory information, [9] a calculation of revenue and expenses, [30] information on the balance sheets, [31] statement of the organization's three largest service accomplishments, [32] and information on organizational structure. [33] In 2010, there were 370,902 Form 990-EZ filers, 60,821 of which filed electronically. [34]

The Form 990-N (also known as the "e-Postcard") is the newest form, created as a result of the Pension Protection Act of 2006 (the "PPA"), which required smaller organizations to file electronic forms (just as the largest organizations are also required to file the 990 electronically). The PPA legislated that these smaller organizations, which were previously not required to file at all, should begin filing annual returns no later than the 2010 tax year to retain their tax-exempt status. [35] The IRS requires organizations that receive less than $50,000 in income [36] to file the Form 990-N electronically. [37] Should an organization decide it wants to file a paper form, it has to complete an entire Form 990. Every organization must provide its employer identification number, tax year, legal name, secondary name, address, officer names and address, and web site, as well as confirmation that gross receipts were less than $50,000 [38] and a confirmation of whether the organization is active or not. [39] The only time a paper form is required of an organization that files a Form 990-N is when there is a change to the organization's "name, structure, or operations." [40] In 2010, there were 449,770 Form 990-N filers. [41]

The Form 990 presents a unique set of challenges for the IRS, and differs from other filings the agency processes. The IRS collects no fees from Form 990 filers, and some of the information included in the Form 990 does not assist the IRS in its mandate as the nation's tax administrator. The IRS does charge an initial user fee when an organization files for tax-exempt status through Form 1023. The organization is required to pay a $400 fee if it has gross receipts less than $10,000 a year over a four-year period, and an $850 fee if it has gross receipts greater than $10,000 a year over a four-year period. [42]

The IRS has not imposed any fees on organizations that submit the Forms 990, even though there has been an increase in the number of nonprofits that file annual returns. In a statement before the House Committee on Ways and Means in 2005, David M. Walker, comptroller general of the U.S. Government Accountability Office, recognized that IRS caseload had increased and recommended "a full re-examination of the tax-exempt sector in light of the challenges facing the nation in the 21st century." [43]

Legal Track:

The kinds of disclosures required in the Forms 990 have changed over the years. The types of questions asked of nonprofits and types of data collected have increased significantly since the first iteration of the Form 990. The most recent legislative changes to the Form 990 occurred in 2008, but the Form 990 has a much longer history that reaches back to the Revenue Act of 1943. The Revenue Act created the requirement for nonprofits to file annual returns. [44] Originally, filings required considerably less information than they do now. Among the disclosures were gross income, receipts, and disbursements. [45] The legislation exempted certain organizations from filing returns as well. This ended with the first iteration of the Form 990. [46]

As the nonprofit community grew, Congress was called to regulate transactions conducted by nonprofits that were arguably outside the mandate of the nonprofits' original mission. These transactions, such as owning department stores or selling real estate, were determined to be "unrelated business income." [47] Unrelated business income is now reported under the Form 990-T. Under the Revenue Act of 1950, Congress legislated that any unrelated business income earned by nonprofits would be taxed as well. Granting an exemption for a stream of business that would compete with for-profit business would create an "unfair competitive advantage." [48] Foundations would not be covered by legislation until a number of years later, after it was determined that a number of foundations were engaged in misconduct. [49] The Treasury Department created a report that outlined a number of legislative strategies, but some in Congress argued that the suggested action was not strong enough. [50] The Tax Reform Act of 1969 provided a statutory definition for private foundations and made them a separate entity from nonprofit organizations. [51]

The Pension Protection Act of 2006 (PPA), while not directly about the nonprofit community, brought about a number of changes to the sector. The PPA required most organizations that earned $25,000 or less in income to file an electronic annual notice, which resulted in the creation of the Form 990-N. [52] In testimony before the House Committee on Ways and Means, George K. Yin, chief of staff of the Joint Committee on Taxation, recommended a proposal that would "require small charitable organizations to file short annual returns" because larger returns would be overly burdensome to small nonprofits. [53] The PPA also added a number of new nonprofit types to the IRC [54] and created a new enforcement tool that would allow for the revocation of tax-exempt status for any organization that failed to file a return for three consecutive years. [55] Prior to this, nonprofits that failed to file their returns would only be fined. [56]

After the enactment of the PPA, Senate Finance Committee member Charles Grassley stated, "Both the House and Senate tax writing committees have expressed interest in the scope of charity and whether § 501(c)(3) organizations are providing a sufficient public benefit by conducting charitable programs commensurate in scope with their resources." [57] This sentiment is reflected as Congress continues to hold discussions, most recently with the Oversight Subcommittee of the House Ways and Means Committee holding hearings on regulatory issues and reporting structures of public charities.

Over the years it became apparent that the Form 990 was no longer capturing sufficient information to properly reflect the growing nonprofit community. In 2008, the Form 990 underwent its first major redesign since 1979. [58] The need for the redesign came from "the increasing size, diversity, and complexity of the exempt sector," and changes were based on the need for "enhancing transparency, promoting tax compliance, and minimizing the burden on the filing organization." [59]

The IRS released a draft of the form for comment and received over 700 responses. After taking these responses into account, the IRS released a second draft and received an additional 120 comments. [60] This led to the final iteration of the Form 990, which is the one that is currently used. This new Form 990 adjusted how compensation for officers was reported, added specific governance questions, such as whether there is a policy in place for potential conflicts of interest, and added new financial information requests.

The most recent version of the Form 990 includes questions on foreign investments for organizations with investments valued at $100,000 or more. This is a change from previous iterations that only requested foreign investment information if the organization had revenues or expenses attributable to foreign investments that were greater than $10,000. Other changes include adjustments to the balance sheet portion and activities involving joint ventures.

APPENDIX 3: CURRENT SOURCES OF FORM 990 DATA

The IRS makes the images of the filed Forms 990 available to the public, but the IRS isn't the sole provider of information surrounding the Forms 990. The nonprofit community and a number of third parties use these publicly available files to produce tools and research for stakeholders in the nonprofit community. In addition to the IRS, which provides some tools itself, the major players in the field are GuideStar, the Urban Institute, the Johns Hopkins Center for Civil Society Studies, and the Foundation Center, but a number of other organizations provide essential services for the nonprofit community and the public.

THE IRS:

The IRS provides a search tool called Exempt Organizations Select Check. Users are able to check on a nonprofit's federal tax-exempt status and filings and can search nonprofits by their EIN number, name, or location. Only Form 990-N filers, nonprofits able to receive tax-deductible contributions, or nonprofits that had their tax-exempt status revoked or are on notice for three years may be searched. You must know which category the organization you are looking for falls under prior to your search. The search function is fairly rudimentary and does not consistently provide helpful results.

GUIDESTAR:

GuideStar is a nonprofit organization that provides information and research on other nonprofit organizations, and it has one of the most well regarded single entity search capabilities on nonprofits. This service involves hosting digitized Forms 990 for various nonprofits. Forms less than three years old can be downloaded one at a time at no cost to the user. Nonprofit organizations are able to update the site on their actions and volunteer further information on their activities. In addition, GuideStar provides information for nonprofit compliance and tools for users to compare and analyze nonprofit organizations. GuideStar offers a large collection of information on individual nonprofits free of charge, including brief financial snapshots, nonprofit governance listings, and the Forms 990 themselves. GuideStar does charge a fee for certain data services, including analysis or compensation reports.

Another service GuideStar offers, in conjunction with BBB Wise Giving Alliance and the Independent Sector, is "Charting Impact." Charting Impact is a tool that can help a nonprofit tell its story. Based on how a participating institution answers five questions, GuideStar creates a report for the nonprofit that summarizes the nonprofit's goals and progress in achieving those goals. Another GuideStar tool is Financial SCAN, developed in conjunction with the Nonprofit Finance Fund, which translates financial information into a report and provides a view of a nonprofit organization's financial health. The service does not cover Form 990-EZ or Form 990-PF filers, and it charges a fee.

Philanthropedia is a service acquired by GuideStar that provides a rating system for nonprofit organizations based on crowd-sourced surveys of philanthropy experts, including professionals in the field, researchers, and nonprofit officers. Participants are not allowed to rate their own institutions. Nonprofits are rated according to perceptions and evidence of the impact they make. Nonprofit listings are organized by sector and can also be found using single-entity search.

TECHNICAL TRACK FOR GUIDESTAR:

The data the IRS makes available to the public are provided unedited, and so may contain errors. Because of this, GuideStar must go through a lengthy process to make Form 990 data usable for its services and research. GuideStar receives scanned Form 990 images on CD/DVDs approximately once a month. They are sent to a third-party digitizer that converts the TIFF files into PDFs. This is a crucial part of the process because the TIFF files treat each page as its own file, meaning a 30-page document is 30 separate TIFF files, as opposed to one PDF. These PDFs are made available on a server that the GuideStar data operations team can access. These PDFs are then linked to the GuideStar site.

The third-party digitizers then continue their work and filter out forms that, according to GuideStar criteria, need to be keyed. The purpose of keying returns is to make the information within them computable. As the returns are keyed, the digitizers do quality assurance tests to make certain that entries match and the lines are added correctly. Situations may arise where GuideStar finds that information is mathematically incorrect or is not in the appropriate location. In these cases, the information will be left unchanged due to the difficulty in determining which portion is incorrect. Once the digitizers validate the keyed information, the XML files are delivered to GuideStar and made available for the various services offered. This entire process, from receipt of the TIFF files to full availability of data on the site, takes approximately two months.

Once users are registered with the site, which is free to the user, they can view returns for the three most recent years of the nonprofit. GuideStar does redact returns, sometimes for a fee, if there is potentially sensitive information still available on the form. Sometimes GuideStar also corrects obvious errors, such as spelling mistakes or errors in dates, but it will not correct any computing errors made by the organization.

Over the past few years, GuideStar has had to go through a labor-intensive process that involved correcting data missing entirely from certain sections, or information incorrectly entered in a section or part. [61] During the IRS intake process, XML data from e-filed returns is made to repopulate a form to look like the Form 990 through the use of a style sheet. This digital conversion produced a number of errors, which were reflected in the TIFF files sent to GuideStar. Errors included values that were transposed, incorrect, or missing. As a result, many of the 2008 forms contain a warning of these potential mistakes. GuideStar went through the process of correcting many of these errors and has seen a decrease in the number of these kinds of errors over the past few years.

The Urban Institute:

The Urban Institute is a research institution that, among other projects, conducts research and provides analysis on the nonprofit community. The Urban Institute's Center on Nonprofits and Philanthropy has a number of projects that focus on providing various tools and levels of analysis on nonprofit organizations. Its core program, the National Center for Charitable Statistics (NCCS), provides nonprofit data and interactive tools to researchers and the public. Its archives of IRS nonprofit data include both current data and historical data back to the 1980s, and its research includes a mix of publications like the Nonprofit Almanac and periodic briefs and reports on special topics.

Its Community Platform Project combines its cleaned and classified IRS data along with census and other data and collaboration tools in an open and highly customized website for use by state and local partners, including foundations, United Ways, universities, and governments. [62] NCCS currently maintains 20 sites around the country.

In addition to the various tools provided to nonprofit community stakeholders, the NCCS provides online tools nonprofits can use to fill out their Forms 990 and 990-EZs and file them electronically with the IRS and various state charity offices, most of which require submission of a Form 990 or 990-EZ as well as a state form each year from the charities subject to their jurisdiction. NCCS also works closely with a number of state charity offices to enable e-filing of their registration and renewal forms. It also maintains the primary e-postcard site which approximately half a million nonprofit organizations use each year for filing these forms with the IRS.

In providing nonprofit data, the NCCS conducts a data-cleaning process similar to GuideStar. NCCS receives data from different streams and posts all of it online. It receives raw data directly from the IRS after it has already provided a portion of the valuable keypunching. This effort is conducted by the IRS Statistics of Income Division, largely for internal purposes. The information is not always accurate, so NCCS conducts a series of quality assurance checks on all records it receives. Between 2,500 and 3,000 of the largest records that fail the checks are then manually reviewed and edited to correct errors, most of which can be resolved by reviewing the full Form 990 images to which they have access. In addition, all organizations are assigned National Taxonomy of Exempt Entities (NTEE) codes so researchers and the general public can easily compare, for example, the numbers and financial health of nonprofit job training organizations by state, county, or zip code.

NCCS also keypunches an average of 60,000 returns each year to capture supplemental information such as the

description of organization purposes and programs; more detailed information on revenues, expenses, asset, and liabilities; governance information; and a range of other items. Some of these data are used in the community platforms; others are used for specific NCCS research projects (such as its periodic reports on international development and environmental organizations); and others are keyed by special request for researchers, state nonprofit associations and others conducting their own research. NCCS prices these custom data sets to cover its direct costs and a small fraction of its overhead.

All of its files are documented and made available on the NCCS website in a number of formats that can be readily used by researchers or others wanting to explore the raw data in greater detail. These files can be used to examine the current status of the nonprofit sector or to examine trends in the sector over the past 15 to 30 years depending on the data series. These data files are placed on servers, in a process similar to GuideStar, and are used by NCCS for analyses and comparison between nonprofit organizations and as the basis for the easy-to-use Table Wizard/Report Builder for the general public or the more sophisticated NCCS DataWeb, which is intended to meet the needs of the research community. The DataWeb gives users greater data analysis variety in their search capabilities by generating custom reports and statistics, mashing up data files, and providing access to organization records. Data are pulled from a number of different file types, and users are able to create their own analyses of varying complexity. The public can access information on individual organizations and use the Table Wizards for free. Researchers are typically charged a modest fee for access to full data files (generally in the $50 to $150 range).

The NCCS offers a number of tools for nonprofits to file their annual returns. Among these tools is the 990 on-line system, which allows users to fill out their return and file either electronically or via physical mail. Nonprofits that take advantage of the service are able to access support staff who will help them with their form, convert the form to PDF if necessary, and use the other tools embedded in the filing process. This was formerly a free service, but nonprofits may now be required to pay, depending on their gross receipts. Any nonprofit that has gross receipts equaling less than $100,000 may still use the service for free, but anything larger than $100,000 in gross receipts falls under a tiered pay structure. In addition, nonprofits that qualify for the Form 990-N are provided their own specific portal for filing their electronic forms. Users are able to put the requested information into the NCCS-provided form and submit it using the portal as well.

The Foundation Center:

The Foundation Center analyzes and compiles data on global grantmakers and the kinds of grants they provide. The Center also hosts five library and learning centers in the United States and over 450 funding information centers worldwide. In each of its regional locations, the Center offers educational programs. The database maintained by the Center houses information on "nearly 100,000 foundations, corporate donors, and grantmaking public charities in the U.S. and 2.1 million of their recent grants." [63] In addition, the Center offers a number of free tools and information on its site. The Center also conducts research on the trends in grantmaking for the nonprofit community.

One of these tools is the Foundation Directory Online, which provides a profile of the 10,000 largest private and community nonprofits. Depending on the user's subscription level, the number of organizations and kinds of information accessible through the database will vary. The basic profile includes information such as geographic focus, financial data, types of support, the organization's EIN number, and program areas. The tool also allows for chart and graph visualizations of an organization's grantmaking. Each has a certain level of granularity with its search capabilities and offers multiple options for defining the parameters of what is displayed. Foundation Directory Online is available free of charge, but only certain organizations are included in the basic package. The editorial team will review information from 35 different streams, including direct grantmaker communication, to ensure they have the most accurate information possible. [64]

The Johns Hopkins University Center for Civil Society Studies:

The Johns Hopkins University Center for Civil Society Studies (CCS) conducts research and provides educational programs on methods of collaboration and innovation between civil society, government, and business sectors. The CCS has a number of nonprofit-based projects, including the Comparative Nonprofit Sector Project,

Nonprofit Economic Data Project, and Nonprofit Listening Post Project. The Comparative Sector Project provides data and analyses for local nonprofit communities in a number of countries. The Nonprofit Economic Data Project researches the US nonprofit community, including volunteer, employment, and financial analyses. The Nonprofit Listening Post Project collaborates with US nonprofits to provide information on trends, strategies, and challenges facing the nonprofit community.

Other Sources of Information:

Give Well functions as a charity evaluator. It uses financial qualifiers for its analysis as well as assessing the impact each nonprofit's program has on its constituents. The site is geared towards individual donors, rather than foundation donors, and promotes what it considers to be the charities most worth donating to.

Great Nonprofits acts as a ratings and review system for nonprofits, but does not rate or review nonprofits themselves. Donors, volunteers, and consumers all provide their experience with a given nonprofit organization. The site partners with other nonprofit-oriented services, such as GuideStar, which allows them to repost any reviews on partner sites to the Great Nonprofits site. There are over 1.2 million nonprofits that may be reviewed on the site.

The University of Pennsylvania Center for High Impact Philanthropy provides a series of tools geared towards nonprofit donors. The UPenn Center offers seminars and workshops for donors and online resources and reports. In addition, a "Donor Toolkit" service is offered. The current toolkit focuses on child health and survival and offers users three donation strategies centered on how they would wish to help.

APPENDIX 4: State Requirements

Highlights:

- Nonprofits are generally required to incorporate at the state level to gain state tax-exempt status and to lobby in the state.

- States differ in their approaches to nonprofit regulation; approximately 40 have state charity offices charged with regulating fundraising and use of charitable assets. Almost all require nonprofits that meet state filing requirements to submit a copy of their Form 990 along with their annual state forms.

- States are at the forefront in actively regulating and prosecuting nonprofit violations.

- The Revised Model Nonprofit Corporation Act is model legislation that gives states a framework by which to establish their nonprofit legislation. The model legislation has been adopted in whole or in part by a number of states.

Analysis:

The federal government is tasked with making certain nonprofits comply with their exemption requirements, but states are predominantly charged with investigating nonprofit misfeasance. There are a number of structures in place to allow for state regulators to oversee nonprofit transactions. The first tool in oversight is requiring nonprofit corporations to seek corporate status by filing at the state level. [65] Incorporation offers two key benefits to a nonprofit: an ability to apply for tax-exempt status and limiting the organization's liability. It is only after an organization has formed a state-recognized nonprofit corporation that it may apply for federal tax-exempt status. To receive tax-exempt status, the purpose set out within the articles of incorporation must focus corporate assets for an approved exempt purpose, among other things. [66] If the articles are filed with the appropriate legislative requirements, the secretary of state or corresponding official issues a certificate of incorporation. [67] After being approved for incorporation, the organization may apply for federal tax-exempt status and exemption from state property, sales, or use taxes. [68]

The federal government has maintained a number of regulations in place, but originally states did not have a coherent, unified strategy for tackling granular regulatory issues. The Model Nonprofit Corporation Act (the "MNCA"), and later the Revised Model Nonprofit Corporation Act (the "RMNCA"), provided states a framework for their nonprofit legislation. The MNCA was drafted so that each state could choose between adopting the framework wholly or adapting it around their needs. A number of states, including Wisconsin, Alabama, North Carolina, Virginia, Nebraska, North Dakota, Oregon, Texas, and the District of Columbia, adopted the MNCA as originally drafted. [69] States continued to adapt their own nonprofit laws around their specific needs, but it was a number of years before any major revisions were made in the MNCA. [70] It took a culmination of other states embarking on new legislative paths for nonprofit regulations, an increase in need for accountability, and a lack of official comments in the MNCA that led the Committee on Nonprofit Corporations to revise the MNCA. [71]

The group began the redrafting process by creating a drafting committee, made up of "practitioners, academics, and government officials." [72] The committee was initially concerned with how new legislation would affect the current nonprofit sector, the contours of what was going to be considered a nonprofit, and the issue of religious nonprofits. [73] The committee was also conflicted by the need to facilitate nonprofit accountability without burdening small, often informal organizations.

In 1986, the committee circulated a draft for comment among members of the community. The draft courted some controversy among experts in the field. These controversies were aimed at what were perceived to be deficiencies in the draft. [74] Areas of protest included classification of religious corporations, the lack of a definition of a "nonprofit corporation", and the powers granted to attorneys general, with many comments affecting the final version of the RMNCA.

The RMNCA differed from its predecessor in a number of areas, but none so much as the requirements it placed on nonprofit officers and executives. The RMNCA attempted to replicate the 1984 Model Business Corporation Act, and included a number of legal standards applicable to directors, and created a higher standard of loyalty, but it also took a number of cues from California nonprofit laws. [75] These cues included breaking up nonprofits into the "public benefit," "mutual benefit," and "religious corporation" classifications. [76] New corporations were given the ability to determine what classification they were applying for. Corporations in existence are classified by specific regulations within the RMNCA. [77] This version also addressed concerns regarding meeting requirements between board members. After the RMNCA was completed, many states later repealed prior laws or adopted the RMNCA in part or whole. There is also an ongoing effort to create a multistate nonprofit registration portal, but its progress is unclear.

Legal Track:

Forming a nonprofit corporation affords a number of protections, but it requires meeting a number of requirements at the state level. Generally, a person forming a nonprofit corporation, known often as an incorporator, must meet these filing requirements with the Department of State or a similar official or agency at the state where they wish to incorporate. Usually, an incorporator must file the entity's articles of incorporation first. To be successful, the articles of incorporation must meet the minimum filing requirements, which are determined by the state. Usually, states will require the name of the corporation, confirmation that the entity qualifies as a nonprofit, the purpose for which the organization has formed, the location of the organization, as well as the names and contact information of directors. The RMNCA requires the corporate name; a statement on whether it is a public benefit, mutual benefit, or religious corporation; address of the initial registered office and initial agent at the office; name and address of each incorporation; statement on whether it will have members; and any provision not inconsistent with law regarding the distribution of assets on dissolution. [78] Proper filings grant the incorporator a certificate of incorporation, which affords them the protection of being a corporation. These protections include reducing liability for officers and directors for losses at the corporate level.

The RMNCA provided new measures for nonprofit accountability that give state regulators a number of tools in holding nonprofit officers and management accountable. A ground floor for record-keeping is established, and records of meetings and member transactions are required, similar to that of the for-profit sector. In addition, the nonprofit is required to provide the secretary of state with the most recent annual report. [79] Nonprofit corporations are also required to disclose financial statements, along with a statement from a public accountant or chief financial officer that attests to the accounting being done following accepted accounting principles and, if applicable, not explains the reason for inconsistencies in previous years. [80] The RMNCA requires nonprofits to file annual reports, which are public documents. [81] While it is common practice for states to require the federal Form 990 filed by the nonprofit to accompany the state annual return, it is not inclusive within the RMNCA. [82] In addition, some states require nonprofits to file annual financial reports.

Organizations that solicit financial contributions within a state are generally required to provide more information than those that do not. [83] This includes specific donations provided for specific purposes, which are usually considered separate from other organizational expenses. Other requirements include reports of corporate assets; [84] annual activities reports; [85] annual registration reports, [86] and sometimes a copy of the filed Forms 990.

Attorneys general are given differing standards in supervising various corporation types, having broader power against some and more limited power against others. [87] The attorneys general must protect the public interest and are empowered to require notice for all significant corporate transactions, which includes indemnification, the transfer of substantial assets, merger, or dissolution. [88] States normally do not oversee religious corporations, which prevents states from requiring annual returns from them. [89]

Appendix 5: Electronic Filing

Highlights:

- A number of organizations are required to file one of the Forms 990 electronically, but the thresholds established require predominantly small or large organizations to do so. That leaves a number of medium-sized organizations with the choice between filing electronically or through a paper form.

- A substantial amount of Form 990 data is displayed online, including a number of sites that offer the public the ability to search for images and sometimes limited data from Forms 990. However, there is no congressional mandate that requires they be placed online. Rather, they must be made available for inspection.

Analysis:

Congress mandated that the Forms 990 of the previous three years be made available by the filing organization, [90] but there is no requirement to make the forms available online. This creates an impediment to easily using the associated data. Organizations are simply required to make their applications for tax-exempt status and subsequent materials, including the Forms 990, available to the public for inspection upon request. [91] Individuals are able to submit a request to inspect these materials either through the organization or with the IRS.

The IRS provides nonprofits a work-around for the inspection requirement. If an organization makes its Form 990 and associated materials "widely available," then it is not required to answer specific requests for disclosure. [92] Options for making the materials widely available include the organization posting a form on its web page or to an online database maintained by a separate entity. [93] To be considered widely available, the web page must inform readers of the document's availability and provide download instructions, be posted in a manner that doesn't alter the original file upon viewing or downloading it, with the exception of any information that is statutorily permitted to be withheld, and not require additional software to view or payment to view or download. [94] Ultimately, Congress did not require that the forms be made available on the Internet, but it is a matter of convenience to do so, say in partnership with entities such as GuideStar. Display services offered by these entities have created an opportunity to both fill a need within the community and allow nonprofits to avoid fulfilling individual disclosure requests.

Originally, the IRS accepted the Form 990 electronically on a voluntary basis, but later expanded to accept the Form 990-PF, too. [95] Outside of certain exceptions, each form may be filed electronically, but only the Form 990-N is exclusively filed electronically. There was push back from the nonprofit community about requiring a form to be filed completely and exclusively online. The Department of Treasury replied by indicating that a paper alternative to filing the Form 990-N is to fully complete (not simply fill in the respective 990-N information) the paper Form 990 or Form 990-EZ. [96] There is no reason to believe a Form 990-N qualifying nonprofit would want to do this, as this only serves to increase the organization's workload.

A study conducted by the Urban Institute found that of the more than 714,000 nonprofits required to file the Form 990-N, approximately 429,000 completed the form. [97] It is unclear why so many organizations failed to file this mandatory electronic form. This information represents the smallest nonprofits, but they constitute a large proportion of all nonprofits. A small, unpublished study by the Urban Institute that sought to track down a sample of approximately 100 nonfilers [98] as well as the overall lack of response by organizations that lost their exempt status, suggests that the vast majority of these organizations had in fact ceased operations.

Mandatory electronic filing now includes organizations other than those that file the Form 990-N. The Internal Revenue Service Restructuring and Reform Act of 1998 ("Reform Act") first recognized the growing trend towards e-filing methods. Senator Robert Byrd said, of the need for electronic filing, "the advent of electric filing technology cannot be ignored as we seek to find ways to make the IRS more responsive to the American taxpayer." [99] The Reform Act amended the IRC to "promote electronic filing and provide incentives towards that end." [100]

Organizations that file the Form 990 or Form 990-PF and file "at least 250 returns in a calendar year, including

income, excise, employment tax and information returns," [101] and with $10 million or more in assets are required to file electronically. [102] Organizations are able to file waivers to exclude them from the mandatory electronic filing on the basis of suffering from technological constraints or undue financial burdens associated with compliance. [103] In addition, the IRS is explicitly not allowed to require electronic filings for individuals, estates, and trusts. [104]

Under current law, the secretary of the treasury is given certain latitude to prescribe electronic filing forms. This power is limited, as was previously mentioned, by an inability to require electronic filing for individuals, estates, and trusts. [105] In addition, the secretary may not require "any person" to file their return electronically unless they file at least 250 returns during the calendar year [106] and must take into account the taxpayer's ability to comply, at reasonable cost, with the regulation.[107] The secretary is able to "promote the benefits of and encourage the use of electronic tax administration programs," [108] and may provide incentives for electronically filed returns. [109]

APPENDIX 6: BREAKDOWN OF THE FORM 990

This section summarizes the Form 990. A brief description is provided of each section, including the kind of information it requests from an organization and in what format. Additional information includes the number of fields, the type of information input, and what page the section is located on.

For purposes of data extraction, it may be helpful to determine which sections hold the most value and the difficulty of performing optical character recognition processes on the most valuable sections. This is a subjective determination, but our recommendation for the sections of the most value include the first page, Part 3, and Parts 7–10. The first page includes information that identifies the organization. Part 3 reflects the organization's mission and its three largest service programs. Parts 7–10 include information on officer compensation, assets, expenses, and a calculated balance sheet. These sections provide information that will allow nonprofits to compare their organization with how other organizations are structured and managed. This information could be used to look at how nonprofits have been affected by the economic recession. It will also provide information that may benefit regulators.

FIRST PAGE DIRECTORY INFORMATION:

- This section includes basic information, such as the organization's name, address, telephone number, gross receipts, and organization type.

- Includes 36 fields that may be checked or require a numerical value input or specific text-based information. Located on page 1.

PART 1. SUMMARY:

- This section focuses on the mission and activities conducted by the organization, governance and structure, information on the prior and current year's revenues and expenses, and total net assets.

- Includes 31 fields that may be checked or require a numerical value input or specific text-based information. Located on page 1.

PART 2. SIGNATURE BLOCK:

- This is where the preparer or officer executes the form.

- Includes 15 fields that may be checked or require a numerical value input or specific text-based information. Located on page 1.

PART 3. STATEMENT OF PROGRAM SERVICE ACCOMPLISHMENTS:

- This section focuses on the organization's mission, any significant changes in transactions since the previous return was filed, and information on the three largest service programs offered by the organization. The service descriptions call for the costs and revenue associated with the service.

- Includes 61 fields that may be checked or require a numerical value input or specific text-based information. Located on page 2.

PART 4. CHECKLIST OF REQUIRED SCHEDULES:

- This section focuses on various questions on financial disclosures and informs the organization what schedule it may have to file.

- Includes 106 fields but consists solely of boxes that may be checked "Yes" or "No." Located on pages 3 and 4.

Part 5. Statements Regarding Other IRS Filings and Tax Compliance:

- This asks the organization more specific information on revenues and the method of receipt. Questions may target specific types of 501 organizations.

- Includes 61 fields, but consists primarily of boxes that may be checked "Yes" or "No." Located on page 5.

Part 6. Governance, Management, and Disclosure:

- This requests information on the methods of governance of the nonprofit, including information on local chapters of the organization, what states the form has been filed in, and makeup of voting members or officers.

- Includes 52 fields that consist of primarily boxes that may be checked "Yes" or "No." Located on page 6.

Part 7. Compensation of Officers, Directors, Trustees, Key Employees, Highest Compensated Employees, and Independent Contractors:

- Organizations are required to provide financial information on organizational officers and employees. The form asks for their name, their title/employment type, average number of hours worked, and reportable compensation from the organization. Additional information requested is information on independent contractors.

- Includes 316 fields that may be checked or require a numerical value input or specific text-based information. Located on pages 7 and 8.

Part 8. Statement of Revenue:

- This requests specific information from various revenue streams, including service programs, grants, and other or miscellaneous forms of revenue.

- Includes 117 fields that may be checked or require a numerical value input or specific text-based information. Located on page 9.

Part 9. Statement of Functional Expenses:

- This requires extensive information on expenses from 501(c)(3) and 501(c)(4) organizations. It breaks down expenses between total, program service, management/general, and fundraising expenses. It looks at how an organization classifies its various needs.

- Includes 140 fields that may be checked or require a numerical value input or specific text-based information. Located on page 10.

Part 10. Balance Sheets:

- This is a balance sheet requiring information from the beginning and end of the fiscal year. It looks at assets, liabilities, and a calculation of any net funds.

- Includes 72 fields that may be checked or require a numerical value input or specific text-based information. Located on page 11.

Part 11. Reconciliation of Assets:

- This is a further calculation of total and net revenue and expenses.

- Includes 7 fields that may be checked or require a numerical value input or specific text-based information. Located on page 12.

PART 12. FINANCIAL STATEMENTS AND REPORTING:

- Information requested regards the method of accounting and the relationship between the preparer and organization.

- Includes 18 fields that may be checked or require a numerical value input or specific text-based information. Located on page 12.

APPENDIX 7: LIST OF INTERVIEWS

1. Telephone interview with Sophie Raseman, Director of Smart Disclosure and Senior Policy Adviser at the U.S. Department of Treasury (February 29, 2012).

2. Telephone interview with Brewster Kahle, Internet Archive (March 1, 2012).

3. Telephone interview with Carl Malamud, President and Founder, Public.Resource.org (March 1, 2012).

4. Telephone interview with Jon Orwant, Engineering Manager at Google (March 6, 2012).

5. Telephone interview with David Blaszkowsky, Senior Policy Advisor at the Office of Financial Research at the U.S. Department of the Treasury (March 12, 2012).

6. Telephone interview with Richard Marker, Wise Philanthropy (March 12, 2012).

7. Telephone interview with Lois Lerner, Director of Exempt Organizations Division, and David Fish, Group Manager of Exempt Organizations Tech Division, at the Internal Revenue Service (March 23, 2012).

8. Interview with Jeff Ubois, Program Officer at the MacArthur Foundation, (March 23, 2012).

9. Telephone interview with Darin McKeever, Deputy Director at the Bill & Melinda Gates Foundation (March 24, 2012).

10. Telephone interview with Ruth Madrigal, Attorney Adviser at the Office of Tax Policy in the U.S. Department of the Treasury (March 29, 2012).

11. Telephone interview with Frank Baitman, Chief Information Officer at the U.S. Department of Health and Human Services (April 6, 2012).

12. Telephone interview with Katie Roeger, Research Associate and Assistant Director at the National Center for Charitable Statistics at the Urban Institute (April 12, 2012).

13. Telephone interview with Cindy Lott, Senior Counsel to the National State Attorneys General Program at Columbia Law School (April 16, 2012).

14. In Person interview with Lois Lerner, Director of Exempt Organizations Division at the Internal Revenue Service (April 23, 2012).

15. Telephone interview with Jennifer Deger, Controller at the Bill & Melinda Gates Foundation (April 17, 2012).

16. Telephone interview with Tom Pollak, Senior Research Associate and Program Director for the National Center for Charitable Statistics at the Urban Institute (April 19, 2012).

17. Telephone interview with Elizabeth Boris, Executive Director at the Center on Nonprofits and Philanthropy at the Urban Institute (April 27, 2012).

18. Telephone interview with Sean Delaney, Executive Director of the Lawyers Alliance for New York (April 27, 2012).

19. Telephone interview with Jacob Harold, former Program Officer at the William and Flora Hewlett Foundation (April 29, 2012).

20. Telephone interview with Chuck McLean, Vice President of Research at GuideStar (May 07, 2012).

21. Telephone interview with Emmy So, Database Operations Assistant at the Foundation Center (May 10, 2012).

22. Telephone interview with Lucy Bernholz, Visiting Scholar at Stanford University Center on Philanthropy and Civil Society (May 11, 2012).

23. Telephone interview with Jon Dunford at the Urban Institute (May 11, 2012).

24. Telephone interview with Jonathan Greenblatt, Director of the White House Office of Social Innovation and Civic Participation (May 11, 2012).

25. Telephone interview with Mike Wash, Chief Information Officer at the National Archives (May 18, 2012).

26. Telephone interview with Bradford Smith, President of the Foundation Center (May 29, 2012).

27. Telephone interview with Hugh Jones, Supervising Deputy Attorney General Tax and Charities Division at the Hawaii Department of the Attorney General (May 29, 2012).

28. Telephone interview with Christine Aube, Director of Data Management at GuideStar (May 29, 2012).

29. Telephone interview with Matt Hampel, Ben Sheldon, and Serena Wales, Code for America Fellows (May 30, 2012).

30. Telephone interview with Peter Swords, (June 4, 2012).

31. Telephone interview with Stacy Palmer, *Chronicle of Philanthropy* (June 7. 2012).

32. Telephone interview with Bob Ottenhoff, former President and CEO of GuideStar (June 6, 2012).

33. Telephone interview with David Fish, Group Manager of Exempt Organizations Tech Division, at the Internal Revenue Service (June 15, 2012).

34. Telephone interview with Bob Carlson, Assistant Attorney General at the Missouri Attorney General's Office (June 27, 2012).

35. Telephone interview with Lin Smith, Managing Director of the National Economics and Statistics Tax Sampling group at PricewaterhouseCoopers (June 28, 2012).

36. Telephone interview with Karen Gano, Assistant Attorney General at the Connecticut *office* of Attorney General (June 29, 2012).

37. Telephone interview with Elizabeth Korsmo, Assistant Attorney General at the *office* of New Mexico Attorney General (July 03, 2012).

38. Telephone interview with Susan Boehmer, Director, and Barry Johnson, Chief Special Studies Branch, of the Statistics of Income Division of the Internal Revenue Service (October 22, 2012).

1. Internal Revenue Service, Office of Research, Forecasting and Data Analysis, Publication 6186 (2011)
2. Molly F. Sherlock and Jane G. Gravelle, An Overview of the Nonprofit and Charitable Sector, Congressional Research Service (2009)
3. Internal Revenue Code Sections 501(c)(1)–501(c)(29); Erika K. Lunder, Frequently Asked Questions About Tax-Exempt Organizations, CRS Report 96-264; Erika K. Lunder, Characteristics of and Reporting Requirements for Selected Tax-Exempt Organizations, CRS Report RL30877
4. Internal Revenue Code Section 501(c)(3)
5. Quick Facts about Nonprofits, NCCS.Urban.org, http://nccs.urban.org/database/overview.cfm#core (last visited)
6. Virginia G. Richardson and John F. Reilly, Public Charity or Private Foundation Status Issues under IRC 509(a)(1)-(4), 4842(j)(3), and 507, 2003 Exempt Organizations-Technical CPE
7. IRC Section 509(a)
8. Molly F. Sherlock and Jane G. Gravelle, An Overview of the Nonprofit and Charitable Sector, Congressional Research Service (2009)
9. IRS Business Master File, July 2009, Urban Institute National Center for Charitable Statistics
10. Molly F. Sherlock and Jane G. Gravelle, An Overview of the Nonprofit and Charitable Sector, Congressional Research Service (2009)
11. Molly F. Sherlock and Jane G. Gravelle, An Overview of the Nonprofit and Charitable Sector, Congressional Research Service (2009)
12. Molly F. Sherlock and Jane G. Gravelle, An Overview of the Nonprofit and Charitable Sector, Congressional Research Service (2009)
13. Molly F. Sherlock and Jane G. Gravelle, An Overview of the Nonprofit and Charitable Sector, Congressional Research Service (2009); Wing, Pollak, and Blackwood, The Nonprofit Almanac, 2008, P. 134 and CRS calculations
14. Molly F. Sherlock and Jane G. Gravelle, An Overview of the Nonprofit and Charitable Sector, Congressional Research Service (2009), Page 38
15. IRM Section 4.75.10.4-1(A)
16. IRM Section 4.75.10.4-1(B)
17. IRM Section 4.75.10.4-1(C)
18. IRM Section 4.75.10.4-1(D)
19. Form 990, Lines B-M
20. Form 990, Part 1 Lines 8-22
21. Form 990, Part 6, Lines 1a-9
22. Form 990, Part 3, Lines 1–4e
23. Office of Research, Forecasting and Data Analysis, Internal Revenue Service, 2011 Publication 6186
24. Form 990-PF, Lines A – J
25. Form 990-PF, Part 1 Lines 1-27c
26. Form 990-PF, Part 2, Lines 1-31
27. Form 990-PF, Part 4, Lines 1-3
28. Office of Research, Forecasting and Data Analysis, Internal Revenue Service, 2011 Publication 6186
29. Form 990-EZ, Lines 1-L
30. Form 990-EZ, Part 1 Lines 1-21
31. Form 990-EZ, Part 2 Lines 22-27
32. Form 990-EZ, Part 3 Lines 28-32
33. Form 990-EZ, Part 4
34. Office of Research, Forecasting and Data Analysis, Internal Revenue Service, 2011 Publication 6186
35. Frederick J. Gerhart, Charitable Incentives and Charitable Reforms under the Pension Protection Act of 2006,
36. http://www.irs.gov/Charities-&-Non-Profits/Annual-Electronic-Filing-Requirement-for-Small-Exempt-Organizations--Form-990-N-%28e-Postcard%29
37. IRC Section 6033(i)(1)
38. Katie L. Roeger, Small Nonprofit Organizations A Profile of Form 990-N Filers, Charting Civil Society (Urb. Inst., D.C.), Aug. 2010
39. IRC Section 6033(i)(2)
40. https://www.independentsector.org/form_990
41. Katie L. Roeger, "Small Nonprofit Organizations A Profile of Form 990-N Filers, Charting Civil Society" (Urb. Inst., D.C.), Aug. 2010
42. http://www.irs.gov/charities/article/0,,id=139480,00.html; Form 1023, Part 11
43. Capitol Hill Hearing Testimony, House Ways and Means, April 20, 2005; http://www.gao.gov/assets/120/111547.html
44. William H. Byrnes, "The Private Foundation's Topsy Turvy Road in the American Political Process," 4 Hous. Bus. & Tax L.J. 496 (2004)
45. Revenue Act of 1943 Section 117(a), 26 U.S.C. Section 54 (1944)
46. Paul Arnsberger, Mellissa Ludlum, Margaret Riley, and Mark Stanton, A History of the Tax-Exempt Sector: An SOI Perspective (citation?); also see, Chasin et al. Form 990. http://www.irs.gov/pub/irs-tege/eotopicg02.pdf
47. Diane L. Fahey, Taxing Nonprofits Out of Business, 62 Wash. & Lee L. Rev. 547 (2005)
48. Paul Arnsberger, Mellissa Ludlum, Margaret Riley, and Mark Stanton, A History of the Tax-Exempt Sector: An SOI Perspective
49. 1965 Treasury Report
50. Thomas A. Troyer, The 1969 Private Foundation Law: Historical Perspective on Its Origins and Underpinnings, the Exempt Organization Tax Review Jan. 2000 Vol. 27 No. 1
51. Tax Advisory Group of the Office of Legal and Regulatory Affairs, American Health Lawyers Association, The Public Charity Status of Hospitals and Health Care organizations Under the "Operated in Connection With" Relationship Test of Section 509(a)(3) of the Internal Revenue Code of 1986, 22 AHA. Jour. Of Health Law (1989)
52. IRC Section 6033(i)
53. George K. Yin, "An Overview of the Tax-Exempt Sector," 109 CIS Legis. Hist. P.L. 280 (109 PL 280) (2005)
54. U.S.C. Section 1223(a) (2006)
55. IRC Section 6033(j)
56. Frederick J. Gerhart, "Charitable Incentives and Charitable Reforms Under the Pension Protection Act of 2006," 19 Health Law. 17 3 (2007)

57. Evelyn A. Lewis, "Charitable Waste: Consideration of a "Waste Not, Want Not" Tax," 30 Va. Tax Rev. 39 (2010)
58. Karen Donnelly, "Good Governance: Has the IRS Usurped the Business Judgment of Tax-Exempt Organizations in the Name of Transparency and Accountability?," 79 UMKC L. Rev. 163 (2010)
59. Summary of the Form 990 Redesign Process,
60. Id.
61. IRS, Known DVD Image Issues for Electronically Filed Returns (Form 990 Family), as of December 30, 2011
62. See SHARE New Mexico, at http://www.shareNM.org, or http://www.theNonprofitLink.org (Jacksonville) for a couple of examples.
63. http://foundationcenter.org/about/;jsessionid=J35WDLRXVQHQ5LAQBQ4CGXD5AAAACI2F
64. https://secure1.foundationcenter.org/fdo/signup/faq
65. Nicole S. Dandridge, AALS 2011 Annual Meeting, Section on Nonprofit Law and Philanthropy: Choking Out Local Community Service Organizations: Rising Federal Tax Regulation and Its Impact on Small Nonprofit Entities, 695 K.Y. L.J. (2010)
66. NPOLT 2:2
67. Mod Nonprofit Corp Act § 31 (1964)
68. Stephen D. Simpson, 2008 Multistate Guide to Regulation and Taxation of Nonprofits, (CCH 2007)
69. NPOLT 1:12
70. Oleck, Mixtures of Profit and Nonprofit Corporation Purposes and Operations, 16 N. KY. L. Rev. 225, 238 (1989)
71. Moody
72. Moody
73. Minutes of the Meeting of the Committee on Nonprofit Corporations (Nov. 9, 10, 1979)
74. Moody; Revised Act, supra note 29, at 3:01(a)
75. NPOLT Section 1:11
76. Cal. Corp. Code Sections 5110-6910, 7110-8910, and 9110-9690 (1989)
77. Revised Act at Section 17.07 (1) – (5)
78. NPOLT 2:2
79. Revised Act at Section 16.01
80. Revised Act at Section 16.02
81. Revised Act at Supra Note 29, Section 16.22
82. Office of the New York State Attorney General, Charities Bureau Form CHAR 023; N.J. Stat. Ann. Section 45:17A-24
83. Dana Shilling, Lawyer's Desk Book, Section 1.15 (2012)
84. Cal. Gov't Code Section 12586(a); N.Y. Est. Powers and Trusts Law Sect. 8-1.4(f)(1)
85. RMNCA Section 7.01(d)(1)
86. Charles Nave, "Exempt Organizations Law: Charitable State Registration and the Dormant Commerce Clause," 31 Wm. Mitchell L. Rev. 227, 229, 242 (2004)
87. Revised Act at Section 6.30(f)
88. Id. at §§ 6.30, 14.21, 14.31, 14.02, 14.03, 8.55, 11.02, 8.10, 12.02 and 3.04.[RECHECK]
89. I.R.C. Section 6033(a)(3)(A)(i) (2006); Treas. Reg. Section 1.6033-2(g)(1)(i)-(ii) (2010); I.R.S., U.S. Dep't of the Treasury, Pub. No. 557, Tax-Exempt Status for Your Organization 22 (2010), available at http:// www.irs.gov/pub/irs-pdf/p557.pdf.
90. IRC Section 6104(b)
91. IRC Section 6104(a)(1)(A)
92.   IRC Section 6104(d)(4); 26 C.F.R. Sect. 301.6104(d)-2-2(a)
93. 26 C.F.R. Sect. 301.6104(d)-2(b)(2)(i)
94. 26 C.F.R. Sect. 301.6104(d)-2(b)(2)(iii); Bruce R. Hopkins, The Law of Tax-Exempt Organizations 953, 9th ed. (2007)
95. Capitol Hill Hearing Testimony, House Ways and Means, April 20, 2005
96. T.D. 9454, 2009-2 C.B. 178; Nicole S. Dandridge, AALS 2011 Annual Meeting, Section on Nonprofit Law and Philanthropy: Choking Out Local Community Service Organizations: Rising Federal Tax Regualtion and Its Impact on Small Nonprofit Entities, 697 K.Y. L. J. (2010)
97. Katie L. Roeger, Small Nonprofit Organizations A Profile of Form 990-N Filers, Charting Civil Society (Urb. Inst., D.C.), Aug. 2010
98. Personal communication from Tom Pollak (The Urban Institute).
99. 144 Cong. Rec. 4505 (1998)
100. Anthony D. Skidmore, The Internal Revenue Service is Shifting American Taxpayers From a Paper–Based Filing System to Electronic Filing–Is the IRS Offering a Capable System That Protects Taxpayer Confidentiality?, 367 J. Marshall J. Computer & Info. L. (2001)
101. IRS Releases Discussion Draft of Redesigned Form 990 for Tax-Exempt Organizations, IRS.gov, http://www.irs.gov/newsroom/article/0,,id=171329,00.html (last visited)
102. http://www.irs.gov/efile/article/0,,id=108211,00.html; IRC Section 6033(i)
103. IRC Section 6033(i)
104. IRC Section 6011(e)(1)
105. IRC Section 6011(e)(1)
106. IRC Section 6011(e)(2)(A)
107. IRC Section 6011(e)(2)(B)
108. IRC Section 6011(f)(1)
109. IRC Section 6011(f)(2)

THE ASPEN INSTITUTE

One Dupont Circle, NW
Suite 700
Washington, DC 20036
www.aspeninstitute.org