

# The Successful Effort to Open Up Non-Profit Data and its Implications for Government Data

**Nathan Dietz**

**Senior Researcher, Do Good Institute**

**School of Public Policy – University of Maryland**



SCHOOL OF  
PUBLIC POLICY  
**DO GOOD INSTITUTE**

XBRL US Municipal Data Forum  
Washington, DC: April 24, 2019

# Nonprofit Filing Requirements



	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.



# IRS-SOI EO Data Products



- EO Business Master File (BMF) Extract (IRS)
- EO Annual Extracts (IRS-SOI)
- EO Sample Studies (SOI)



# EO BMF Extract



- Cumulative listing
- Includes organizations with active exemption, regardless of Form 990/990-EZ or 990-PF filing requirement
- All 501(c) subsection codes
- Records: 1,702,638 (as of 4/8/2019)
- Only 28 fields; 3 are financial



# EO Annual Extracts (RTF – Return Transaction Files)



- Include selected organization-level data for population of returns from EO Master File
- All subsection codes
- Minimal data cleaning to correct major transcription and/or arithmetic errors
- Processing year file (doesn't correspond to tax year)



# EO Financial Extracts



- Processing Year 2012 file was first one made widely available to the public
- For Processing Year 2013, the number of fields was significantly expanded

Form	Records	Fields	% Financial
990	289,603	245	61%
990-EZ	218,981	71	80%
990-PF	100,484	179	71%
<b>Total</b>	<b>609,068</b>	<b>495</b>	<b>68%</b>



# EO Sample Studies



- “Stratified” random sample – sampling rates

Size of assets (\$ M = millions)	Sampling rate	Number of organizations selected for sample
<b>Very Large</b> >\$50m	100%	9,034
<b>Large</b> \$20m - \$50m	39% - 100%	6,033
<b>Medium</b> \$5m - \$20m	7% - 100%	7,334
<b>Small</b> \$1m - \$5m	3% - 14%	4,404
<b>Very Small</b> <\$1m	1% - 5%	4,829
Total		31,634





# Nonprofit Data Available through NCCS



IRS data releases:

- BMF (multiple versions / year)

- SOI Sample and Annual RTF Extracts (annually)

NCCS-GuideStar National Nonprofit Research Database (“digitized data”) (1998-2003)

NCCS Core Files (1989-present)

NCCS Master File (with NTEE codes)

NCCS Trend Analysis Files



# Part III: Program Service Accomplishments



Form 990 (2013)

Page **2**

**Part III** **Statement of Program Service Accomplishments**  
Check if Schedule O contains a response or note to any line in this Part III ☐

**1** Briefly describe the organization's mission:

**2** Did the organization undertake any significant program services during the year which were not listed on the prior Form 990 or 990-EZ? ☐ Yes ☐ No  
If "Yes," describe these new services on Schedule O.

**3** Did the organization cease conducting, or make significant changes in how it conducts, any program services? ☐ Yes ☐ No  
If "Yes," describe these changes on Schedule O.

# Part III, Continued: Details of Program Service Accomplishments



4 Describe the organization's program service accomplishments for each of its three largest program services, as measured by expenses. Section 501(c)(3) and 501(c)(4) organizations are required to report the amount of grants and allocations to others, the total expenses, and revenue, if any, for each program service reported.

4a (Code: ) (Expenses \$ including grants of \$ ) (Revenue \$ )

4b (Code: ) (Expenses \$ including grants of \$ ) (Revenue \$ )





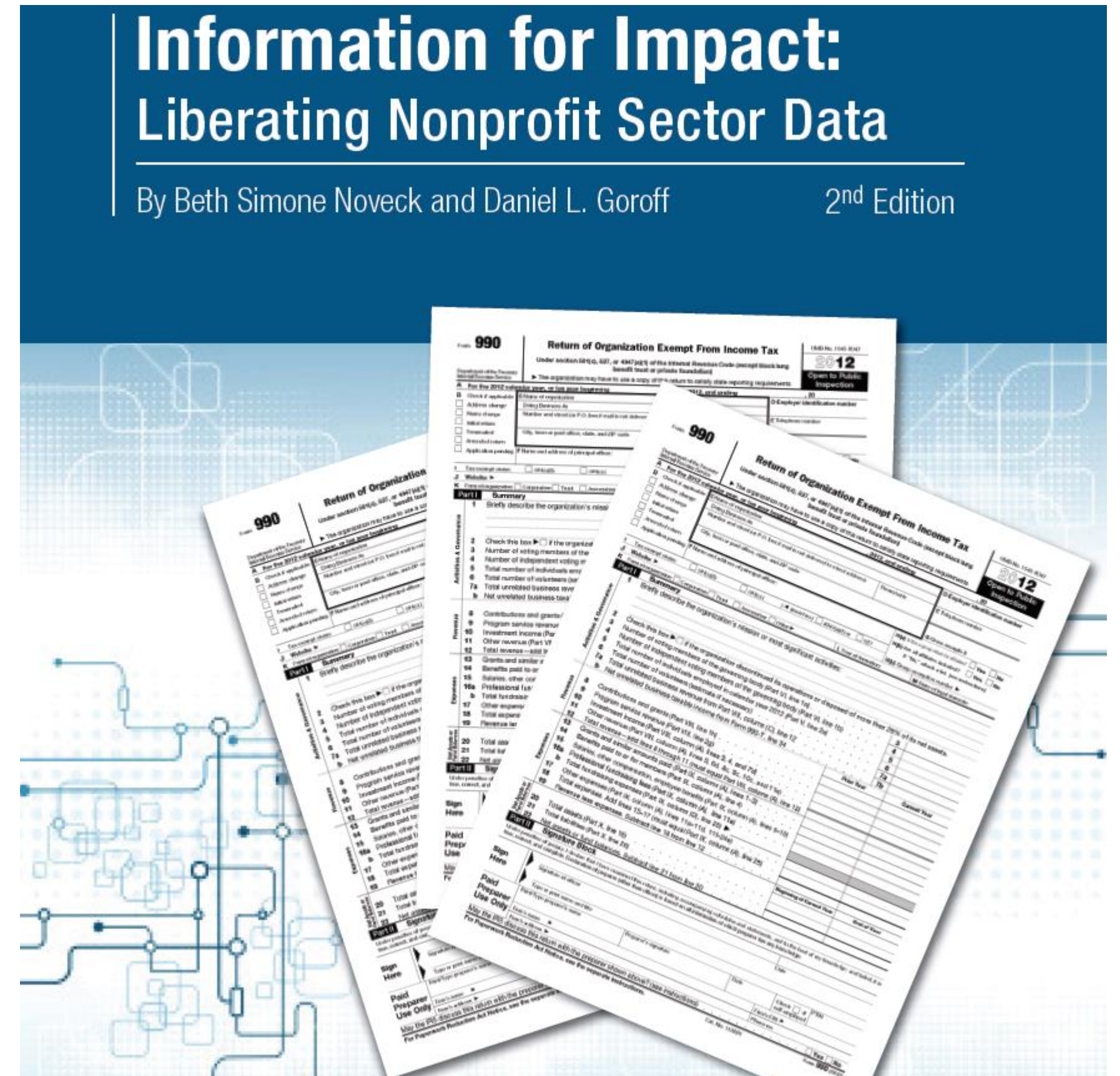
# Highlights of 990 Open Data Initiative



- 2013: Aspen Institute report, *Information for Impact*, published
- 2015: PublicResource.org wins case against IRS
- 2016: IRS releases data from e-filed returns in XML format
- 2017: Aspen Institute hosts first “Validatathon” to process XML 990 and 990-EZ data
- 2017: Nonprofit Open Data Collective formed
- 2019: H.R. 1957, Taxpayer First Act of 2019 , passes the House



SCHOOL OF  
PUBLIC POLICY  
DO GOOD INSTITUTE



# Uses of Open 990 Data



- Locate expertise and resources
- Aggregate and compare programmatic expenditure trends
- Identify leaders and their connections
- Aggregate and compare specific income trends
- Promote transparency and trust
- Assess financial soundness
- Track and compare labor statistics
- Check investment and program alignment
- Aid enforcement and regulation
- Monitor lobbying
- Facilitate data “mashups”
- Facilitate academic and business research



# Open 990 Data Resources



- IRS AWS documentation page: <https://docs.opendata.aws/irs-990/readme.html>
- *Charity Navigator resources: 990 Decoder* (<https://github.com/CharityNavigator/irs990>)  
*and 990\_long* ([https://github.com/CharityNavigator/990\\_long](https://github.com/CharityNavigator/990_long))
- IRSx: <https://github.com/jsfenfen/990-xml-reader>
- Nonprofit Open Data Collective GitHub site: <https://nonprofit-open-data-collective.github.io/overview/>
- Master Concordance File: <https://github.com/Nonprofit-Open-Data-Collective/irs-efile-master-concordance-file>
- Open990.org: <https://www.open990.com/org/> - flat files available

“The 990 dataset is nothing but hierarchical data. In 2018, developers use JSON for hierarchical data, and store JSON in NoSQL databases.” – David Borenstein





# Remaining Issues:



- Processing issues (requires DIY solutions):
  - Separate *parent* Xpaths for each item in a list
  - Same parent, different child Xpath for each item in a list
  - Two “lines” for the same field
  - Two different Xpaths for the same thing
  - Text in ALL CAPS
  - A field whose definition is different on the paper form
- Communications with IRS EO division
  - Missing schemas (a recent problem)
- **Historical data from paper returns**



# Additional Resources



Aspen Institute Program on Philanthropy and Social Innovation:

<https://www.aspeninstitute.org/programs/program-on-philanthropy-and-social-innovation-psi/>

*Information for Impact* (2013 report):

[https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/Information for Impact Report FINAL REPORT 9-26-13.pdf?\\_ga=2.83780822.1728374377.1556632822-1769392819.1556632822](https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/Information for Impact Report FINAL REPORT 9-26-13.pdf?_ga=2.83780822.1728374377.1556632822-1769392819.1556632822)

David Borenstein article on processing XML data: <https://medium.com/@open990/the-irs-990-e-file-dataset-getting-to-the-chocolatey-center-of-data-deliciousness-90f66097a600>

David Borenstein interview about 990 data:

<https://aws.amazon.com/blogs/publicsector/qa-open990-and-how-to-access-financial-indicators-from-nonprofits/>



# For Questions and More Information...



**Nathan Dietz**

*pronouns: he/him/his*

**Senior Researcher, Do Good Institute**

University of Maryland | School of Public Policy  
2107 Susquehanna Hall, College Park, MD 20742

Office: (301) 405-0207 | Cell: (202) 650-7599

[ndietz@umd.edu](mailto:ndietz@umd.edu) | [dogood.umd.edu](http://dogood.umd.edu)



SCHOOL OF  
PUBLIC POLICY

DO GOOD INSTITUTE