

Beginner’s Guide to Using NCCS Data

Introduction

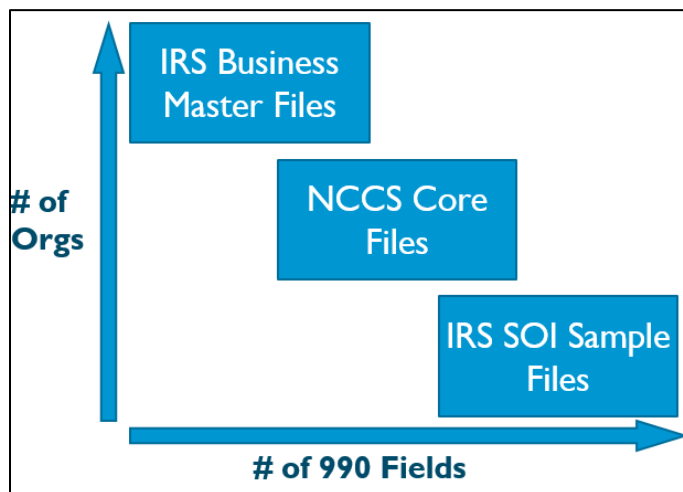
This guide presents an overview of best practices and general guidance for using National Center for Charitable Statistics data. All data referenced in this guide is freely available on the [NCCS Data Archive](#). All files are presented in CSV file format for easy download and access.

For a brief overview of the key databases available on the NCCS Data Archive, as well as advice on choosing a particular file for analysis, see the section “Key Files,” below. Once a file is chosen, users should refer to the section “Key Filters” for best practices on applying standard filters to the data. Users are strongly encouraged to review these sections before conducting analysis using NCCS data.

Key Files

The three most widely used datasets hosted by the NCCS Data Archive are: the IRS Business Master Files (BMF); the NCCS Core Files (Core); and the IRS Statistics of Income Exempt Organizations Sample Files (SOI). Each dataset has different advantages and limitations that should be considered before choosing a file for analysis.

Roughly speaking, each dataset operates on a tradeoff between the number of organizations covered by the dataset and the number of fields available in the data. See below:



IRS Business Master Files (BMF): BMF datasets are generally released multiple times per year; at least one per year is available going back to 1995, with many years (especially the most recent) containing multiple files. Files are [updated by the IRS on a regular basis](#), and downloaded to the NCCS Data Archive several times per year. Each file represents the number of organizations **registered** with the IRS during the month/year indicated; the information in the BMF is almost always the most up to date information about these organizations and their filing statuses.

As a comprehensive **list of all registered organizations**, the BMF contains records both for active and inactive organizations, as well as registered organizations operating below the minimum requirements for filing a Form 990 or 990-EZ.

However, while the BMF does contain recent information on geographic location and type of organization, the number of total fields from the Form 990 are *very* limited. Financial fields are limited solely to total revenue and total assets, and many Parts of the Form 990 are not represented at all.

- **Typical use:** the BMF is often used to get the most recent available counts of numbers of nonprofit organizations registered or active in the US, or in a particular geographic area. For tracking raw numbers, in number of new organizations or estimating closures.
- **Key considerations:** the IRS purports to remove registered organizations from the BMF after three consecutive years of inactivity. In practice, however, this process can take longer. If researchers seek to get lists of the number of **active** registered organizations, they are highly advised to filter the BMF to records where CFILER = "Y" and CZFILER = "N". This will ensure that the data is restricted to nonprofits filing within the last two years of the release of the BMF (CFILER = "Y") *and* filing with more than \$0 in gross receipts (aka, not a 990-N e-postcard filing; CZFILER = "N").

NCCS Core Files: Released once per year, the NCCS Core Files represent the entire population of **active, reporting** organizations of the given type filing within a given **circa year**. The NCCS Core Files are separated into three sets, based on type of organization: PC (Public Charity), PF (Private Foundation) and CO (all Other 501(c) organizations). **Only organizations that file a Form 990, 990-EZ or 990-PF are counted as active and reporting.**

Each Core file contains the most recent filing available for a given organization within two years of the Core File date. Thus, each Core File contains multiple fiscal years' worth of information, but only one record for each EIN. This data is aggregated to show an accurate snapshot of actual nonprofit activity within the given year, accounting for the varying filing patterns and fiscal years of different nonprofits.

The NCCS Core Files contain a far wider selection of financial fields from the Form 990 than the BMF, including detailed information on revenue, expenses, and assets. Further, the Core files are subject to the NCCS validation procedure before release, making the data more accurate than that contained in the BMF.

The information used to create the NCCS Core files are generally older than available for the BMF. Further, the Core files only contain information for organizations filing a Form 990, 990-EZ, or 990-PF, not all registered organizations as in the case of the BMF.

- **Typical use:** the most popular of the NCCS data sets, and the one with the broadest application for nonprofit analysts, the breadth of fields available make the NCCS core files useful for many types of analysis, including the state of the nonprofit sector in any given year, detailed financial and/or growth information in a particular geography, etc.
- **Key considerations:** for more fields than available in the NCCS Core Files, users should consider the NCCS Core Full files, which contain a larger number of fields. These files are currently available for 2012-2015. However, users should note that these files pertain **only** to organizations filing the “full” Form 990, and thus exclude organizations that file a Form 990-EZ (which would be covered by the general NCCS Core files).
 - Researchers interested in longitudinal analysis should consider using the NCCS Core Fiscal Year trend file. Aggregated by merging all available NCCS Core Files through 2013, the NCCS Core Fiscal Year Trend file contains only one record per EIN per fiscal year. Thus, while separate examination of two consecutive Core Files may contain the same information for a given organization (if the organization’s next filing was not available as of the release of the second file), the Core Fiscal Year Trend file would only include one listing. This makes the file very suitable for longitudinal analysis of particular organizations.

IRS Statistics of Income Exempt Organizations Sample Files (SOI): The SOI files are **weighted sample** files of organizations filing a Form 990 or 990-EZ. The files take a proportionate sample of all organizations filing with the IRS, weighted to include the largest organizations and nonprofit hospitals. The data in these files are thoroughly checked and verified by the IRS before release (as opposed to the BMF, which are not as thoroughly checked, or the Core files, which are partially manually verified by NCCS).

These files contain the broadest selection of fields, including many fields not covered in the Core Files. However, this is only for a sample of all nonprofits, and extrapolating for the entire population of nonprofits should be done with care.

- **Typical use:** detailed analysis of fields not covered by the NCCS Core Files.
- **Key considerations:** the SOI files are weighted solely by size of nonprofit organization, as measured by total assets. Thus, all aggregated findings should use the WEIGHT variable to extrapolate proper estimations. Furthermore, any estimations based on other nonprofit indicators (e.g., geography, type, or financial ratios) should be done with extreme care.

Key Filters

When conducting analysis using any of the above data sets, NCCS recommends that users keep in mind the following data preparation steps and filters:

Filter “Out of Scope” organizations:

All main NCCS files contain a field labeled OUTNCCS, the “Out of Scope Flag”. NCCS recommends that all analysis conducted uses only records where OUTNCCS = “IN”. OUTNCCS is a binary flag for whether the organization has been deemed out of scope for US nonprofit sector analysis. The most frequent reason for exclusion is being a foreign-based entity filing with the IRS, or organizations operating in US Territories or overseas. For the reason any particular organization is flagged as “Out of Scope”, please see the field OUTREAS.

Apply current NTEE code

For any analysis conducted on nonprofit type, researchers are advised to use the organization’s National Taxonomy of Exempt Entities (NTEE) code. This 3-5-digit alphanumeric code classifies each nonprofit organization by their primary purpose. Each digit of the NTEE code represents an increasing level of specificity. For example, a nonprofit arts organization specializing in television production and broadcasting would be classified as “A32”. “A” is the common classification for all arts organizations. “A3” corresponds to media and communications organizations. And finally, “A32” stands for Television organizations. For more information on NTEE codes, see <http://nccs.urban.org/classification/national-taxonomy-exempt-entities>.

While most NCCS data files contain fields for NTEE code, NCCS data users are strongly advised to use the codes found in the NCCS “Current Master NTEE Lookup file” (labeled as “nccs.nteedocAllEins.csv”, available here: <http://nccs-data.urban.org/data.php?ds=misc>). This list is the NCCS master list of all the most recent NTEE information available for any particular organization. NTEE codes are subject to change: an organization may change its primary purpose, the IRS and/or NCCS may decide that a different code better fits the organization’s primary purpose, or in some instances a particular organization may note that they have been misclassified and request a change by NCCS. In these instances, NCCS does *not* change all files containing that organization, but rather instead updates the NCCS Master list. Therefore, users are strongly encouraged to match the NCCS Master list against any other files they are using and use the NTEEFINAL field from the master list for any particular organizational EIN.

Filing requirement codes

The BMF and Core files both contain fields for “Reason for 501(c)(3) status” (FNDNCD) and “Filing Requirement Code” (FRCD). Users are advised to take note of these fields to filter out organizations possibly not intended for analysis. For example, many NCCS reports focusing on public charities filter out private foundations (classified as FNDNCD “02”, “03”, or “04”), as well as organizations that are not technically required to file due to their religious status (FRCD classification of “060”, “061”, “130”, or “131”). Users are encouraged to consider these fields when conducting analysis.

Minimum filing threshold

Prior to 2010, the minimum filing threshold for an organization required to file a Form 990-EZ was \$25,000 in gross receipts. In 2010, this minimum was raised to \$50,000 in gross receipts (although there is evidence for a delayed transition period between the two). Depending on research interest, NCCS data users should consider restricting analysis solely to organizations required to file during the time period of interest (i.e., having gross receipts or total revenue above the minimum threshold). Organizations below the filing threshold may choose to file a Form 990-EZ or even the full Form 990. In some instances, this information will be relayed in the IRS information used by NCCS. However, in other instances the IRS will instead enter “0” for every financial field, reflecting the fact that that particular organization falls below the threshold. Unfortunately, for these “zero filers” it is not generally possible to distinguish between organizations that authentically have \$0 in all financial fields and those codes as such by the IRS for falling below the minimum threshold. Therefore, it is generally good practice to restrict financial analysis solely to organizations required to file.

Furthermore, any longitudinal analysis that bridges periods using the earlier filing threshold and the newer filing threshold should keep the change in mind. For example, any analysis comparing the 2009 NCCS Core File to the 2010 NCCS Core File may see more organizations filing in 2009 than in 2010. However, the drop may be attributable to the change in minimum reporting threshold, not to decline in number of organizations.

BMF “C” Fields.

The BMF contains several fields that represent fields cross-validated with the NCCS Core files. These fields are denoted by the “C” prefix. As noted above, the BMF generally does not receive comprehensive validation or verification of particular fields before release by the IRS. However, for the limited financial fields available within the BMF, as well as some other fields, NCCS cross-checks the information available in the BMF with the information in the NCCS Core files. This cross-checked information is presented in these “C” fields. Users are highly advised to defer to the “C” fields over the plain fields when possible. For example, users should employ CTOTREV for “total revenue” rather than TOTREV. The original fields are, however, retained for completeness.

Other Considerations and Resources

NCCS users should keep in mind the limitations of both each data set individually as well as NCCS data in general. The data produced by NCCS are intended for use by researchers and policy-makers in their quantitative analyses, and as a springboard for more in-depth survey or case study research. While NCCS employs a range of data checking procedures, our limited resources do not permit systematic in-depth verification of every record. If your focus is on individual organizations or on categories where a few large organizations account for the majority of financial activity or resources, we strongly advise that you include inspecting financial data for dominant individual organizations as part of your analysis plan.

Users should be aware of organizations not covered by IRS data. For example, NCCS does not provide comprehensive information on organizations not required to register with the IRS, such as very small organizations (those with less than \$5,000 in annual gross receipts) or religious organizations.

Information presented in NCCS data sets may similarly be confounded by subsidiary organizations filing using the same EIN as umbrella organizations. In these instances, only the largest record per EIN is retained in NCCS data set. Therefore, subsidiary organizations may not be wholly captured.



Users interested in detailed specifics of a small subsector or region may want to consider manual review of Form 990 data, such as that available in pdf format using the Foundation Center’s 990 Finder tool: <http://foundationcenter.org/find-funding/990-finder>. Manual review of the Form 990 can reveal information not covered by the aggregated data and is always more accurate than data presented in collected form.

Finally, users should be aware of other nonprofit sector resources available:

- Other research conducted by the Urban Institute’s Center on Nonprofits and Philanthropy: <https://www.urban.org/policy-centers/center-nonprofits-and-philanthropy>
- For information on charitable giving and contributions: [Giving USA](#).
- For information on foundation/grant data: [The Foundation Center](#)
- For other Form 990 information: [Guidestar](#)
- For other e-filed Form 990 information: [Open990](#)