Accessing IRS Form 990 Data with Excel

Humberto Barreto  
*DePauw University, hbarreto@depauw.edu*

Michele T. Villinski  
*DePauw University, mvillinski@depauw.edu*

Follow this and additional works at: https://scholarship.depauw.edu/econ_facpubs

Part of the Economics Commons

Recommended Citation

This article was published as: Barreto, Humberto and Michele T. Villinski. 2019. "Accessing IRS Form 990 Data with Excel." *Philanthropy & Education*, Vol. 2, No. 2 (Spring), pp. 95-110. Published by Indiana University Press. Stable URL: https://www.jstor.org/stable/10.2979/phileduc.2.205 No part of it may be reproduced, stored in a retrieval system, transmitted, or distributed in any form, by any means, electronic, mechanical, photographic, or otherwise, without the prior permission of Indiana University Press. For education reuse, please contact the Copyright Clearance Center. For all other permissions, contact IU Press at ."
Accessing IRS Form 990 Data with Excel

Humberto Barreto
DePauw University

Michele T. Villinski
DePauw University

Form990.xlsm, freely available at archive.org/details/Form990.xlsm, is a standalone, macro-enabled Excel workbook that allows users to easily download and analyze IRS Form 990 tax data. This information is extremely difficult to access in machine-readable format, but the IRS recently made the data available as XML files. Form990.xlsm directly imports data in separate sheets for each institution. This paper demonstrates how to use this tool by analyzing a small sample of similar colleges. We compare Form 990 data with IPEDS and find substantial agreement, but important differences. Form 990 data offer researchers an alternative to IPEDS for higher education work and enable evaluation of hospitals, political associations, service organizations, and other nonprofits.

Keywords: nonprofit, endowment, tuition, IPEDS

The Return of Organization Exempt from Income Tax, known as Form 990, is how nonprofits report financial information to the Internal Revenue Service (IRS). First filed for the 1941 tax year with two pages and three questions (Chasin, Kawecki, & Jones, 2002), Form 990 has become the key source of data on tax-exempt organizations. The U.S. Department of the Treasury reports that

Humberto Barreto, Ph.D. is the Q. G. Noblitt Professor of Economics and Management at DePauw University. He has written several books and many articles on how to use Excel® spreadsheets to improve the teaching and learning of economics and econometrics. He holds an annual teaching workshop for faculty interested in becoming more effective teachers. Visit www.depauw.edu/learn/econexcel for more information.

Michele T. Villinski, Ph.D. is Professor of Economics and Management at DePauw University. Her previous research interests include health care delivery and access, water options pricing models, and integrating technology into economics pedagogy.
“for Tax Year 2014, over 293,000 charities exempt under section 501(c)(3) filed Forms 990 with the IRS reporting over $3.7 trillion in assets and nearly $2.0 billion in revenue” (n.d. b, p. 1).

On June 16, 2016, “the Internal Revenue Service announced that the publicly available data on electronically filed Forms 990 will now be available for the first time in a machine-readable format through Amazon Web Services (AWS)” (U.S. Department of the Treasury, 2016, para. 1). Before widespread computer use, Form 990 data were only available by direct request to the IRS and, more recently, only image files have been supplied. Extracting data was tedious and cumbersome. The AWS release is a significant improvement in openness and accessibility because the data are available as XML files, making it much easier to process information from millions of returns.1

This paper further lowers the costs of access by showing how to use Form990.xlsm, a macro-enabled Excel workbook that offers an easy interface to the AWS data. Form990.xlsm is freely available, with open-source Visual Basic code, at archive.org/details/Form990.xlsm. The workbook contains detailed instructions and documentation.

We used Form990.xlsm to download a few variables on a small sample of private, selective liberal arts colleges. We compared the Form 990 data to the most common source used by higher education researchers, the Integrated Postsecondary Education Data System (IPEDS). We find substantial agreement, but there are important differences for individual observations between Form 990 and IPEDS data.

The next section reviews Form 990 and IPEDS data sources. Section 3 describes our sample and introduces the Form990.xlsm Excel workbook. Results can be found in section 4, followed by our conclusion.

IPEDS AND FORM 990

Financial data used in research on higher educational institutions comes from three main sources: independent/researcher-generated surveys, IPEDS, and the IRS Form 990. Independent surveys are relatively rare, while IPEDS is the most frequently used resource. The National Center for Education Statistics (NCES), part of the U.S. Department of Education and the Institute of Education Sciences (U.S. Department of Education, n.d. b), collects information on a wide range of financial, student, enrollment, and institutional characteristics from over 7,500 U.S. colleges, universities, and vocational and technical schools each year in the IPEDS reporting process. Every institution that participates in federal student financial aid programs must report to IPEDS (U.S. Department of Education, n.d. a). Data are available to researchers, students, and the public online (U.S. Department of Education, n.d. c). Each institution designates a keyholder, the individual responsible for overseeing the IPEDS data collection and entry process. Typically, the keyholder works with offices across the institution to provide data for fall (October), winter (February), and spring (April) collection deadlines (U.S.
Department of Education, 2018). A keyholder described the process and offered the following opinion:

In the decades since the introduction of IPEDS, many improvements have been made in both upload and download processes. IPEDS uses variance formulae to detect data that are inconsistent with previous years, allowing institutions to verify or correct, and provide an explanation for variation outside the norm. Many collection components have file upload options, which also greatly increase efficiency for keyholders. IPEDS takes great care in verifying the reliability of reported data, following up with keyholders when questions arise. This is the primary reason that release of data takes longer than researchers and the public like (W. Tobin, personal communication, September 20, 2018).

IPEDS is an omnipresent data source in research on higher education, in part due to the large number of institutions, variables, and years it covers. For example, Hedrick, Wassell, and Henson (2009) relied solely on IPEDS data in their analysis of trends in higher education administrative costs. They examined 17 years of data for 249 schools. In contrast, Powell, Gilleland, and Pearson (2012) combined IPEDS data on institutional characteristics, expenditures, and retention and graduation rates with Carnegie classifications and variables from the National Study of Postsecondary Faculty to devise a model for assessing both efficiency and effectiveness in higher education.

While IPEDS has been the dominant data source, information on financial aspects of higher education and other nonprofit institutions is also available from annual IRS Forms 990. According to the instructions for filing, “Form 990 must be filed by an organization exempt from income tax under section 501(a) . . . if it has either (1) gross receipts greater than or equal to $200,000 or (2) total assets greater than or equal to $500,000 at the end of the tax year” (U.S. Department of the Treasury, n.d. a, p. 3). For the 2014 tax year, nearly 300,000 tax-exempt organizations filed Form 990 (U.S. Department of the Treasury, n.d. b), including hospitals, colleges and universities, charitable trusts, and political organizations. Charitable organizations that do not file Form 990 include: some churches and religious-affiliated organizations, certain governmental and political organizations (such as the state or local committee of a political party), and specified organizations with gross receipts below the threshold to file ($50,000).

Key categories of data reported on the IRS Form 990 include revenue, expenses, assets, a balance sheet, and names, titles, and compensation of top officers. Researchers should note that, for example, the 2018 Form 990 applies to the fiscal year beginning in 2018; the organization lists the start and end date of the fiscal year in part A of the form. For many higher education institutions, this time period is from July 1, 2017 to June 30, 2018. Compensation information, though, is based on calendar year. IRS Form 990 data are prepared either by the institution’s finance department or by accounting firms hired by the institutions. In our sample, roughly three of four schools used external accountants (35 out of 46 institutions). The Board of Trustees, or a designated subcommittee of the
board, reviews the form before submission. In addition, most schools ensure that financial data and compensation information are reviewed by third parties such as auditors and lawyers. As Hyatt (2008) notes, “The unfortunate institution whose officials cut corners in completing Form 990 risks an IRS audit, penalties, and public embarrassment” (p. 16).

While the IRS Form 990 dataset is rich, it has been underexploited in the higher education literature. This is likely due to practical barriers to accessing the data. Previous options for gathering data from IRS Form 990 were working directly from pdfs solicited from schools, gathered from websites, retrieved from a service such as Guidestar, or paying someone else to harvest the data. These methods are costly in terms of time, price, or both, and they severely limit flexibility in exploring the data. As recently as 2015, Parker used financial data from IRS Form 990 to assess the costs of switching athletic conferences. He accessed data on a select number of variables for 15 institutions via Guidestar, then entered it into Excel before analyzing it in SPSS (Parker, 2015).

Another way Form 990 data have been used is through The Chronicle of Higher Education’s annual compilation of presidential compensation data (e.g., Langbert, 2006). The Chronicle’s current presidential compensation analysis is an online interactive tool that relies on IRS Form 990 data for private schools (Bauman, Davis, & O’Leary, 2018). The Chronicle data are available for $249 per year. However, accessing the IRS Form 990 data directly would yield compensation data for employees beyond the president.

Krishnan, Yetman, and Yetman (2006) compared data from IRS Form 990 with audited financial statements and manual analysis of webpages in their research into fundraising and program expenses in nonprofit organizations. Like Parker, they used Guidestar as a source for IRS Form 990 data. Krishnan et al.’s data collection process was labor intensive: their IRS Form 990 data were not in spreadsheet form and they obtained some years from Guidestar and others from the National Center for Charitable Statistics database.

While these are merely a few of the studies that have used IRS Form 990 data, they illustrate common issues of accessibility that have existed in the past. Form 990 data have been expensive to gather and cumbersome to analyze. The IRS initiative to make Form 990s available in electronic format through the AWS opens opportunities for automated retrieval, quick importation into spreadsheet and statistical software, and vastly increased flexibility in analysis. In the next section, we present a new, straightforward, customizable tool for accessing and manipulating IRS Form 990 data and give a modest demonstration comparing IRS Form 990 to IPEDS for a small sample of liberal arts colleges. The IRS Form 990 dataset may be better for some purposes than IPEDS. The Form 990 dataset encompasses a wider range of institutions and covers some variables not included in IPEDS. According to Capaldi and Abbey (2011), who examined ways universities can better use data to improve their performance, “different universities apply different rules for answering questions on the same [IPEDS] forms” (p. 15). In contrast, Krishnan et al. (2006) concluded that “use of an outside paid accountant
increases reliability and usefulness of nonprofit financial reports” as most schools do when completing their tax forms (p. 399). Differences in the data collection and submission processes for IPEDS and the IRS Form 990 ultimately affect the quality of the financial data available to researchers. One industry insider rated IRS Form 990 as eight out of 10 on quality of data input on financials, with IPEDS earning four out of 10 (B. Kelsheimer, personal communication, October 5, 2017).

METHODS

Our Sample

We began selecting schools and universities to study by downloading data on all 246 schools in the Carnegie basic classification “Baccalaureate Colleges: Arts and Sciences Focus” (The Carnegie Classification of Institutions of Higher Education, n.d. a). We used information from Carnegie on school characteristics to narrow the sample to 53 small (or very small) nonprofit, private, 4-year, selective (or more selective), highly residential, exclusively undergraduate colleges and universities focusing on the arts and sciences. Using IRS Form 990 data, we removed seven additional very small schools with fewer than 800 employees, for a final sample of 46 institutions. Table 1 shows our sample selection procedure.

Table 1. Selecting the Sample

<table>
<thead>
<tr>
<th>Sample size after each filter</th>
<th>Filter Description</th>
</tr>
</thead>
</table>
| 246                           | Carnegie basic classification: Baccalaureate colleges: Arts and sciences focus  
                                    “Includes institutions where baccalaureate or higher degrees represent at least 50 percent of all degrees but where fewer than 50 master’s degrees or 20 doctoral degrees were awarded during the update year.” (The Carnegie Classification of Institutions of Higher Education, n.d. c, para. 5) |
| 245                           | Nonprofit          |
| 218                           | Private            |
| 81                            | Undergraduate profile: Arts and sciences focus—no graduate coexistence  
                                    “At least 80 percent of bachelor’s degree majors were in the arts and sciences, and no graduate degrees were awarded in fields corresponding to undergraduate majors.” (The Carnegie Classification of Institutions of Higher Education, n.d. e, para. 7) |
| 63                            | Graduate program—(not classified)  
                                    *This indicates the school confers exclusively undergraduate degrees.* (The Carnegie Classification of Institutions of Higher Education n.d. b, p. 5) |

(Continued)
Table 1 (Cont.)

<table>
<thead>
<tr>
<th>Sample size after each filter</th>
<th>Filter Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>More selective (or selective), lower transfer-in</td>
</tr>
<tr>
<td></td>
<td>&quot;Fall enrollment data indicate at least 80 percent of undergraduates are enrolled full-time at these bachelor’s or higher degree granting institutions. Test score data for first-year students indicate that these institutions are more selective in admissions (80th to 100th percentile of selectivity among all baccalaureate institutions). Fewer than 20 percent of entering undergraduates are transfer students. Note: 40th to 80th percentile for selective schools. (The Carnegie Classification of Institutions of Higher Education, n.d. f, para. 14, 15)</td>
</tr>
<tr>
<td>53</td>
<td>Small (or very small), highly residential</td>
</tr>
<tr>
<td></td>
<td>&quot;Fall enrollment data indicate FTE enrollment of 1,000 – 2,999 (fewer than 1,000) degree-seeking students at these bachelor’s or higher degree granting institutions. At least half of degree-seeking undergraduates live on campus and at least 80 percent attend full time.&quot; (The Carnegie Classification of Institutions of Higher Education, n.d. d, para. 9, 12)</td>
</tr>
<tr>
<td>46</td>
<td>More than 800 employees</td>
</tr>
<tr>
<td></td>
<td>(Form 990, Part I, Line 5)</td>
</tr>
</tbody>
</table>

The 46 colleges and universities in our sample are distributed across the United States as shown in Figure 1, which displays beginning-of-year endowment.
Form990.xlsx Macro-enabled Excel Workbook

Our macro-enabled Excel workbook is largely self-explanatory, thus, we offer a simple overview and instruction set here. After downloading the file from archive.org/details/Form990.xlsx, open the workbook and make sure that security settings allow macros to run. Review the Intro worksheet for details about software requirements and detailed instructions for using the data retrieval tool. Clicking the EIN tab takes the user to the sheet that actually accesses the data. The first step is to click the “Get CSV” button and enter the tax year under study. The next step is to enter Employer Identification Numbers (EINs) for each tax filer. Figure 2 shows an example.

Use the index sheet in the workbook (downloaded in step 1) to search by institution name to find the tax filer’s EIN. Excel formulas and database functions (e.g., VLOOKUP) can be used in the index sheet to return EINs given institution names. Clicking the “Get XML” button will populate the Taxpayer Name column and download Forms 990 as separate sheets in the workbook.

<table>
<thead>
<tr>
<th>EIN</th>
<th>TAXPAYER NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. The EIN sheet in Form990.xlsx.

Aside from entering or importing the list of EINs, using the tool is a straightforward series of button clicks to:

a) Import tax form index information for a given tax year in CSV format from AWS,

b) enter specific EINs, and

c) download IRS Form 990 data for given EINs into Excel.

While we have focused on a group of small college and universities, these functions of the Form 990 data retrieval tool can be applied to any institution or group of institutions that filed electronically and whose form has been included in the AWS releases. For instance, other users might choose to study hospitals, private foundations, or a different slice of higher education (such as athletics conferences).

For our higher education data, the Form 990 Excel workbook also includes built-in macros for analyzing Form 990 after accessing and downloading it from AWS. Three buttons in the EIN sheet examine data related to endowment, tuition, and compensation. Researchers using the Excel workbook to examine institutions...
RESULTS

Following the instructions above, we used Form990.xlsm to download XML data for our sample institutions for the 2013 tax year (July 1, 2013 to June 30, 2014 for most schools). The filing deadline is May 15, 2015, and the 990 forms appear over the next several months in the AWS database.

We then used the endowment, tuition, and compensation buttons to extract a few variables and saved the results in three separate Excel files available at archive.org/details/Endowment. We checked the XML data with image (pdf) files to confirm they were the same. Finally, we downloaded these same variables from IPEDS and compared the two sources.

We report summary results below, but all of the data are available in the Excel files. It is easy to identify specific schools and sort by different variables.

Endowment

The Excel workbook, Endowment.xlsx, contains all the data and results described in this section. Form 990 asks schools to report the beginning and ending endowment for a five-year period, as shown in Figure 3 for one institution in our sample (Barnard College).

![Figure 3. Example of 2013 Form 990 Endowment data.](image)

We compared the values for beginning of year balance (line 1a, column (a) Current Year) to the IPEDS variable, Value of endowment assets at the beginning of the fiscal year (F1314_F2_RV). We also compared the values for the end of year endowment balance. For Barnard, IPEDS and Form 990 were identical.

We found that 25 schools had exactly or almost exactly the same values in the two data sources, but seven of 46 schools had a greater than 5% difference in
beginning endowment. These same seven schools, with Franklin and Marshall and Amherst having greater than a 10% discrepancy, also had large differences in ending endowment. Figure 4 displays the results for the full sample.

![Comparing Beginning Endowment](image1)

![Comparing Ending Endowment](image2)

**Figure 4.** Comparing Form 990 and IPEDS Measures of Endowment.

**Source:** Comparing sheet in Endowment.xlsx.

We used Form 990 data to compute compound annual growth rates (CAGR) for the endowment over the five-year period and annual endowment draws. The CAGR includes all contributions to the endowment (fund-raising); investment earnings, gains, and losses; grants or scholarships (drawn from the endowment); other expenditures for facilities and programs; and administrative expenses. For Barnard, the CAGR was

\[
CAGR_{\text{Barnard}} = \left( \frac{282,048,169}{165,305,949} \right)^{\frac{1}{5}} - 1 \approx 11.3%.
\]

Figure 5 shows the distribution of CAGRs in our sample. Barnard’s 11.3% annualized return was quite good. The average CAGR for this time period for our 46 schools was about 8.8%. The data show a great deal of dispersion, with two schools (Lycoming and Knox) growing their endowments at extremely fast rates of 14.5% and 16.7%, respectively, while several schools generated CAGRs under 5% per year.

<table>
<thead>
<tr>
<th>Variable</th>
<th>CAGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Obs</td>
<td>46</td>
</tr>
<tr>
<td>Number Non-missing</td>
<td>46</td>
</tr>
<tr>
<td>Number Missing</td>
<td>0</td>
</tr>
<tr>
<td>Number Unique values</td>
<td>46</td>
</tr>
<tr>
<td>Mean</td>
<td>8.78%</td>
</tr>
<tr>
<td>SD</td>
<td>2.69%</td>
</tr>
<tr>
<td>Min</td>
<td>3.20%</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>7.34%</td>
</tr>
<tr>
<td>Median</td>
<td>8.17%</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>10.33%</td>
</tr>
<tr>
<td>Max</td>
<td>16.67%</td>
</tr>
</tbody>
</table>

**Figure 5.** Five-year Endowment CAGRs.

**Source:** HistCAGR sheet in Endowment.xlsx.
Schools also vary widely in their use of endowment draw. Most university administrators want to see the draw below 5% per year and, at many colleges, presentations of the budget highlight the proposed draw for the upcoming year. Form 990 data offer a fresh perspective by enabling computation of the realized, or ex post, draw.

Our annual endowment draw variable computes the grants or scholarships, other expenditures for facilities and programs, and administrative expenses as a fraction of the average of beginning and ending endowment amount that year (See the Doc sheet in Endowment.xlsx for a demonstration that the average endowment provides a good approximation when computing the annual draw). For Barnard, we sum lines d, e, and f in Figure 3 and divide by the average endowment balance for each year. For the current year (2013–14 tax year) for Barnard, the draw was:

\[
\text{Current year endowment draw}_{\text{Barnard}} = \frac{5,741,752 + 4,842,748 + 1,178,328}{240,710,000 + 282,048,169} \approx 4.5\%.
\]

Barnard’s average draw for the five-year period was 4.9%, which is a little better than the average draw in our sample, as shown in Figure 6. Lycoming had the lowest average draw by far, at 1.2%, while eight out of 46 schools had average draws above 6%, with Union College taking the highest average draw of 8.4%.

---

**Figure 6.** Five-year Average Endowment Draws.

*Source: HistDraw sheet in Endowment.xlsx.*

**Tuition and Student Enrollment**

The data and results in this section are available in the Excel workbook *Tuition.xlsx*. Schools report the amount of tuition and fees they billed in Part VIII: Statement of Revenue of their Form 990, as shown in Figure 7 for Dickinson College.
Like Dickinson, most schools separate out various other revenue sources, but some lump all other sources into a single number. Carleton and Centre colleges are the only two institutions in our sample that combined room and board and tuition. Carleton separated out the room and board charge on their web site, and we called Centre to obtain their room and board charge.

Given billed revenue from tuition and fees for each school, we can compute a high-quality estimate of full-time enrollment (FTE) since total revenue equals price X quantity. Dividing billed tuition and fees by the school’s list (or sticker) price for tuition and fees (which is easily available) yields FTE. Figure 8 compares the Form 990 FTE with the FTE reported in IPEDS. While broadly in agreement, there are substantial differences. For all of the schools in our sample, IPEDS reported an FTE count of 82,626, while the Form 990 estimate was 80,318 FTE students. There were 22 out 46 schools with more than a 5% difference in FTEs, led by Bates, Scripps, and Spelman with differences greater than 20%.

Figure 8. Comparing Form 990 and IPEDS Measures of FTE.
Source: Comparing sheet in Tuition.xlsx.
Along with the endowment draw, administrators at small colleges often report and focus on the financial aid fraction, the proportion of list price tuition that is subsidized by grants from the institution. Since colleges report how much grant aid they give students on Form 990 (line 13 in Part 1, on the first page, reports the sum of grants in Part IX, column (A), lines 1–3), dividing this by billed tuition tells us the fraction of tuition subsidized by grants to students. Dickinson reported $40,682,351 in grants to students, which is 36.9% of $110,348,110 in gross billed tuition. This is below 46.5%, the average tuition subsidy in our sample of schools. Figure 9 shows that there is a great deal of variation in how much schools subsidize students, from a low of 27.2% by Scripps to a high of 65.0% by Grinnell.

<table>
<thead>
<tr>
<th>Variable</th>
<th>%Tuition Subsidized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Obs</td>
<td>46</td>
</tr>
<tr>
<td>Number Non-missing</td>
<td>46</td>
</tr>
<tr>
<td>Number Missing</td>
<td>0</td>
</tr>
<tr>
<td>Number of Unique values</td>
<td>46</td>
</tr>
<tr>
<td>Mean</td>
<td>46.5%</td>
</tr>
<tr>
<td>SD</td>
<td>10.3%</td>
</tr>
<tr>
<td>Min</td>
<td>27.2%</td>
</tr>
<tr>
<td>25th Percentile</td>
<td>36.1%</td>
</tr>
<tr>
<td>Median</td>
<td>49.9%</td>
</tr>
<tr>
<td>75th Percentile</td>
<td>54.8%</td>
</tr>
<tr>
<td>Max</td>
<td>65.0%</td>
</tr>
</tbody>
</table>

Figure 9. Fraction of Tuition Subsidized by Endowment Grants to Students.

Source: HistSub sheet in Tuition.xlsx.

Compensation

Form 990 requires listing current officers, directors, and trustees, regardless of compensation, and up to 20 current key employees (IRS). This is the source of reports on compensation for presidents by The Chronicle of Higher Education and news stories on other highly-paid people, such as coaches and athletic directors. The data for our sample exactly matched the Chronicle’s data, available behind a pay wall at www.chronicle.com (Bauman et al., 2018).

While IPEDS does not have presidential salary data, we used our small sample to examine the difference in male-female presidential compensation and explore the ratio of presidential to average employee pay. Download Compensation.xlsx to access the data and results in this section.

The IRS allows organizations to submit compensation data “for the calendar year ending with or within the organization’s tax year” (U.S. Department of the Treasury, n.d. a, p. 26). Thus, if a new president or other key employee comes on board in July at the beginning of the fiscal year (typical in higher education), Form 990 will show what compensation they received for the months worked in the calendar year. Our sample tax returns are for the 2013–14 fiscal year, and organizations are reporting compensation data for the 2013 calendar year.

We removed 10 presidents who started in 2013 and, therefore, received partial year compensation for that calendar year and the president of Holy Cross, who
does not receive a salary. Figure 10 shows that there was little difference in total compensation received by male and female presidents in our sample.

![Histogram of High to Avg Ratio](image)

**Figure 11.** Highest to Average Compensation Ratio.

**Source:** *CompPresFullYear* sheet in *Compensation.xlsx*.

We used the information provided in Form 990 data to measure the ratio of the highest paid person to the average pay. Form 990 reports the number of employees that were given a W-2 and the total compensation paid to all employees so calculating the average compensation is simple (see column H in the *Compensation* sheet). We found the maximum total compensation paid to an individual (usually the president, but not always) at each institution and divided by the average compensation (see column I). Figure 11 shows the wide variation in the high-to-average ratio, with many schools in the teens, but others with the highest paid person receiving 20 and 30 times the average compensation. Bowdoin is the outlier at almost 45.
Like presidential pay, this measure is subject to the vagaries of interim leadership and hiring in the middle of the year. Furthermore, some schools include food and other service workers as their employees, while others outsource. Even with Form 990 data, making good comparisons remains quite difficult.

CONCLUSIONS AND FUTURE RESEARCH

While Form 990s have been available as image (pdf) files for some time, extracting information was tedious and expensive. Thus, the availability of Form 990 data in machine-readable format is a true step forward in transparency and offers a rich, new source of information. These data provide an alternative to IPEDS, currently the most common source for higher education research.

Neither Form 990 nor IPEDS data are perfect. They are produced separately via different methods. IPEDS is a complicated survey managed by a single person (usually in a department of institutional research) coordinating input from various sources (such as admission, finance, and human resources). There are missing data issues and imputation methods are applied. Most researchers download and use IPEDS unquestioningly. Form 990 data, on the other hand, are produced by the business office operating under an accounting culture. The Form 990 information is vetted by third parties according to generally accepted accounting principles, submitted to the IRS, and institutions are aware that they could be audited.

We compared Form 990 data with IPEDS and found some agreement, but substantial individual deviations. Endowments varied by more than 5% in seven of 46 observations. We estimated FTEs based on Form 990 billed tuition data and found a greater than 5% difference from IPEDS in 22 out 46 schools, with a few over 20%.

Form 990 data allow for research into areas not covered by IPEDS. For instance, we computed CAGRs as a measure of endowment performance and endowment draws. Both showed substantial variation in our sample. We also found wide dispersion in the fraction of tuition subsidized by financial aid. Finally, Form 990 data offer the opportunity to study individual compensation. We found no gender difference in presidential pay (replicating exactly The Chronicle of Higher Education’s data which are behind a paywall) and a great deal of variation in high-to-average pay ratios.

Several websites provide Forms 990 as pdfs and there are many software options to read Form 990 XML data (e.g., Python). Open990, freely available at www.open990.com, offers access to Form 990 tax information from the AWS public datasets. This paper presented a macro-enabled Excel workbook, Form990.xlsm, a convenient, customizable, user-friendly tool to quickly get AWS Form 990 information in a familiar spreadsheet form. Simply download and open the workbook, then follow the step-by-step instructions. We used it on a sample of private, selective liberal arts colleges, but Form990.xlsm could gather data for research on other nonprofits, such as hospitals, political associations, and service organizations.
ACKNOWLEDGEMENTS

We thank Frank Howland, Brad Kelsheimer, and Bill Tobin for discussions, suggestions, and corrections, and Scott Parsons for georeferencing. We also thank Hope Sutherlin for data validation.

This research did not receive any specific grant funding from agencies in the public, commercial, or not-for-profit sectors.

Note

1. Visit lecy.github.io/Open-Data-for-Nonprofit-Research for more on the campaign to liberate 990 data.

References


