

Machine-Readable Form 990 Launched!

07.05.16 | Linda J. Rosenthal, JD



On June 16, 2016, the Internal Revenue Service made it happen: the long-awaited introduction of the machine-readable Form 990 (Modernized E-file, “MeF”):

”

...[T]he 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).

Of course, IRS was pushed into this new age of sunshine and transparency, but the agency came around. It worked hard to put this important technological upgrade into effect, although there was a slight delay beyond earlier-announced date of early 2016.

Now that the new system is here, there’s a lot of excitement about it:

”

...[T]hough it takes years to liberate data, when the clouds part, you know it! So we should look at this moment, as the IRS begins to publish electronic nonprofit tax returns online in a machine-readable format on Amazon Web Services, as something of a historic watershed.

“The Internal Revenue Service opened a gusher of information on nonprofits ... by making electronically filed Form 990s available in bulk and in a machine-friendly format.”

Why is this Such a Big Deal?

The Form-990-series information returns that nonprofits must file with the IRS already include data that is “legally required to be publicly released.” So a vast amount of information was already – technically – publicly available, but *searching* this data was difficult. “Accessing the filings generally requires a request from the public, which can include a [Freedom of Information Act] request, and results in more than 40 million pages provided in a non-machine-readable format.”

The data from the Form 990 was only available in image files; the filings were previously made public as PDF documents. This format “requires costly manual entry or imprecise character-recognition technology to extract the data in bulk and make it searchable.”

By contrast, under the new technology, “the XML files contain “structured information that represents the main 990 form, any filed forms and schedules, and other control information describing how the document was filed,” according to the AWS web page.” Individual items on the information return can be extracted “easily and precisely.”

“The change means researchers and charity watchdogs such as GuideStar will no longer have to pick and choose which parts of the forms should be keypunched into their databases,” explains Chuck McLean, a senior research fellow at GuideStar. “There are parts of this information that we have never had electronic access to and I will be really interested to see what some of that data is like.”

Which of the Form 990s Are Included?

Data from each Form 990, Form 990-EZ, and Form 990-PF – and all related schedules – will be available as open, machine-readable data for years beginning 2011. The system will be updated each month.

The Form 990-N (e-postcard) used by certain smaller exempt organizations is not part of this new technological advance, but that information can be accessed through available with this data, but it can be accessed IRS.gov.

Not all information from these newly accessible information returns will be available. “Certain donor information” as well as “certain personally identifiable tax-identification numbers” will be omitted “to prevent the data’s misuse.”

The IRS noted in its June 16th announcement that, already a majority of organizations subject to the information return filing requirement submit them electronically. “Both paper and electronically filed 990 returns will continue to have image files made and these files will continue to be available by DVD.”

Conclusion

“Opening nonprofit tax-return data online will be transformative,” according to Carl Malamud of Public.Resource.org, a prime mover behind the court case that spurred this change. “If you make the data available, you will get innovation.”