#990datathon

Hacking the IRS form 990

David Bruce Borenstein, PhD
Charity Navigator
October 2, 2017
Nonprofit Open Data Collective

http://npo-data-collective.github.io/
990 eFiles: an opportunity for learning
Hundreds of thousands of e-files

Number of eFiles on AWS, by tax period year

- **990**
- **990EZ**
- **990PF**
990s vary tremendously in content

https://charitynavigator.shinyapps.io/990tree/
Many representations of same content

26
Distinct schema versions

3
Versions of the IRS form 990

19
Forms in 990 family

8,235
Distinct XML paths

32%
990PC filings lack schema metadata
990 “Datathon” at Aspen
Where we are now

• 26 schemas mapped into 1 “concordance”

• Simplified .csv dataset now available

http://990.charitynavigator.org/
Where we go from here

• Verify mapping and validate data
  – Another big volunteer operation?

• User-friendly resource for searching data
  – GuideStar, ProPublica

• Use by researchers, journalists...and you!