



**Testimony of  
Daniel Quigg  
Chief Executive Officer  
Public Insight Corporation  
September 20, 2017**

Good morning Chairman Smith, Vice Chair Ryan, and Ranking Member Cera, and members of the committee, I am Dan Quigg, CEO of Public Insight Corporation, a provider of a free public platform ([www.Publicinsightdata.com](http://www.Publicinsightdata.com)) that transforms public data to insight. 360-Public launched in June, 2013 collects data from multiple sources and maximizes the utility from that data. I am also a CPA by background and was part of the original XBRL standards committee that digitized SEC public financials. Finally, I have an extensive background in financial analytics having founded a commercial financial analytics software company.

I am here today to testify in favor of House Bill 3. The second and third generations of Internet media have passed and Big Data is the next emerging area of leveraging data across the web. Numerous applications have been developed that extract data for public consumption. However, there remains extensive, largely untapped value within existing public data stores. This data is collected routinely by federal, state, and local agencies, but is largely unused. The DataOhio initiative takes the right step forward in unlocking the value of public data and setting in motion continuous improvement that will, over time, optimize its utility.

What does not work is another data repository that simply houses datasets. This approach has little or no value as evidenced by its usage. Consider Data.Gov, the U.S. Government repository of data launched in 2009. The site boasts over 85,000 datasets. With a plethora of data like this, you would think it would be one of the most popular sites on the Internet. However, nothing could be further from the truth. It's US ranking according to Alexa is a dismal 25,360. By comparison, Case Western Reserve University, a medium sized university is ranked 11,712. Further, according to Alexa, Data.Gov caters to graduate level research as every other demographic user is below the national average. Why?

The answer is simple – it is how the data is served up. Most people would recognize value from the data if they knew it existed and how to get it. Data served up in datasets is like a library organized by putting piles of books in stacks loosely organized by topic. Imagine a library without the Dewey Decimal System? Users of Data.Gov are expected to somehow search through the datasets looking for a needle piece of information in a haystack. Interestingly, 10 states have adopted the same open data site approach as



the U.S. government despite its apparent lack of interest. Ohio, like these states maintains multiple repositories of data stored in many different sites.

One fundamental problem is one of sheer numbers. Consider a simple school district financial statement. It typically has around 30 financial statement line items to the page but carries a minimum of 3 funds or 90 metrics. Not that big a deal except when you further consider that the result is just one year. To truly get a trend line historical analysis, you probably would want to look back 5 years. This means 450 data metrics ... for one school district! That would give you trending with no benchmarking or comparison. To get benchmarking, you would want to compare the baseline school district to multiple other school districts. A comparison to 10 other districts yields 4,500 numbers.

The current Open Data approach lists each dataset within its own year with no comparability. Without a system, a user would have to manually extract the data and construct their own individual analysis. Finally, the analysis is worthless without some level of visualization that shows the entire data set in its proper context.

DataOhio contemplates a system that maximizes the value of data. Instead of searching for data within a cylinder, it contemplates a funnel approach that lets the user rapidly find the data, the context and the comparisons they need. We believe we have constructed just such a system in 360-Public that is currently available and is rapidly gaining traction among K-12 school districts. It takes five minutes to learn 360-Public and it is already free to the general public.

Another critical component of DataOhio is the development of standards. Without standards, data is not comparable. As an example of this, we extracted staffing and compensation data for Ohio school districts. Our goal was to determine average Full-Time Equivalents (FTE's) for each major function. However, we found multiple different definitions of FTE and we quickly abandoned the effort. For example, we found some school districts with the same person listed as a duplicate FTE because they were a coach and a teacher. Like an omelet with a rotten egg, all it takes is one deviant from a standard and comparability is suspect.

Finally, DataOhio is critical because defines how data is posted consistently. The Federal Open Data standard emphasizes the concept of "machine readable". However, this term must be more fully developed. Is a scanned financial statement machine readable? The final result has to emphasize data elements that can be aggregated and compared. The Comprehensive Annual Financial Report or CAFR represents a goldmine of data that can be analyzed and compared if the data components can be extracted and a common taxonomy developed. We have already developed a baseline financial reporting approach in 360-Public.



Mr. Chairman and Members of the committee thank you for this opportunity to testify and I am happy to answer any questions that you may have.

DRAFT