September 5, 2013

Elizabeth M. Murphy
Secretary
U.S. Securities and Exchange Commission
100 F Street, N.E.
Washington, D.C. 20549-1090

Dear Ms. Murphy,

I am writing to address some of the recommendations of the Investor Advisory Committee regarding the SEC and the need for the cost effective retrieval of information by investors, adopted July 25, 2013,[1] as well as some of the comments to the Concept Release on the U.S. Proxy System (File No. S7-14-10, dated July 22, 2010, subsequently "Release") objecting to tagging Form N-PX.

I am an Associate Professor of Economics at the Stanford Graduate School of Business. One of my research areas is voting by mutual funds in shareholder meetings. I have published two articles on the topic (both joint with Gregor Matvos of Chicago Booth Graduate School of Business), one studying voting on M&A proposals[2] and the other studying voting in director elections.[3] We also have another project on the topic in progress. In the process of writing these articles, I have worked extensively with the N-PX voting data provided by the funds to the SEC (both in raw form in which these forms are available on Edgar, and using processed data from third party service providers).

So I am very familiar with the issues with the data as it is currently submitted and distributed, and with the formidable challenges of working with that data. Prior to my academic career, I have also worked as a software engineer, which gives me a good sense of which issues are easy to address using software tools, and which ones are not.

In this letter, I will not address the question of whether mutual funds (and/or other institutional investors) should be required to disclose their voting decisions. However, given that they already *are* required to disclose these decisions, I feel strongly that the votes should be submitted to the SEC in a standardized, structured, tagged, computer-readable format. **I believe most mutual funds already have their voting data stored in a structured database form, either internally or on a service provider's computers, and use computer scripts to produce the N-PX forms that they then submit to the SEC**. Modifying these scripts so that they produce output in a different format (a structured, standardized one) is straightforward. And having the data in a structured form would turn it from a vast, largely disorganized collection of text files, into a highly useable database.

There are several advantages to having N-PX forms in a structured, computer-readable format. One is that if the data is stored in such a way, then it becomes straightforward to convert it to different formats, enabling all kinds of uses: it would be easy to automatically convert these forms to such formats as spreadsheets for summary analysis, PDF for printing, HTML for web display, plain text for maximum portability across devices, and various database formats for deeper data analysis. In contrast, in the current form, the data is only available in the unstructured form.

[1] http://www.sec.gov/spotlight/investor-advisory-committee-2012/data-tagging-resolution-72513.pdf

[2] "Cross-Ownership, Returns, and Voting in Mergers," *Journal of Financial Economics*, v. 89(3), September 2008, pp. 391-403.

[3] "Heterogeneity and Peer Effects in Mutual Fund Proxy Voting", *Journal of Financial Economics*, v. 98(1), October 2010, pp. 90-112.

If the votes were reported in a structured form (and thus, e.g., could be downloaded in an Excel format), it would be easy for investors to see the overall patterns in funds' behavior, both across funds and over time. For instance, in my 2010 JFE paper with Gregor Matvos, in Table 1, we report the overall level of "management friendliness" (i.e., the fraction of the time a fund votes "For" management-sponsored directors) for 10 large S&P 500 index funds, for two years. The range of management friendliness is very wide, going from Dreyfus, which withheld support from management-sponsored directors only 0.2% of the time in 2003-04, to Vanguard, which withheld support 17.2% of the time, even though these funds invest in roughly the same companies. With structured data, analysis like this would be easy to undertake for any investor or analyst who is familiar with basic spreadsheet software. In contrast, given how N-PX reports are currently submitted, in unstructured plain text formats, undertaking such a project requires considerable time and effort, since before one can compute these simple summary statistics, he or she needs to first reverse-engineer the formats the funds use to report their votes (and different funds do use many different formats, and even the same fund may switch from one format to another from one year to the next), then write and debug computer scripts to extract these votes, and only then it is possible to compute such summary statistics. Similarly, with unstructured data, there is no easy way to see how the voting patterns (e.g., the frequency of supporting management-proposed directors, or opposing shareholder proposals) of a particular fund evolve over time. With tagged data, such questions could be answered virtually instantly.

It is even more challenging and time-consuming to see which funds voted which way on a particular issue in a particular shareholder meeting. For instance, if a controversial merger or shareholder proposal passes by a slim margin, and an investor wants to know which funds voted for it and which voted against, he or she would have to manually open every single one of the thousands of N-PX submissions, formatted in all kinds of different ways, and collect this information fund by fund, which is prohibitively time consuming. In contrast, with structured data downloadable in a spreadsheet or database format, this information could be extracted in a matter of minutes.

In addition to allowing systematic analysis, tagging may spur the creation of websites and services summarizing the data and presenting it to users in a variety of ways, focusing on specific types of issues, allowing the users to "play" with the data and slice and dice it along various dimensions. Right now, creating any such service is a massive undertaking. With tagging, it would turn into a weekend project for any competent web developer.

Some of the comments on the Release question the need for tagged data, arguing that there are already service providers that process N-PX data and make it available to users. This is not a valid objection, for several reasons.

First, commercial service providers charge substantial fees for access to the processed data, which they have to do in order to cover their costs. These fees make the data unaffordable for individual investors.

Second, even though the quality of the data processed by the service providers has been going up over time, it is (unavoidably!) far from perfect. The coverage is not complete, because getting complete coverage would be prohibitively expensive: there are thousands of mutual funds, and they all use all kinds of different formats to report their votes. There are gaps in the data, with data for some funds in years X and X+2, but not X+1. There is no robust way to track the same fund over time, because if the fund changes its name or even records it slightly differently (e.g., "JANUS ADVISER FORTY FUND" vs "JANUS FORTY FUND," to give just one example), service providers'

software interprets it as a new fund. Note that these issues are not the fault of the service providers -- it is just the inherent nature of working with raw, untagged data.

Third, individual fund formats may (and do) change over time, so service providers have to keep spending time and costly effort tracking these changes, writing new software to reflect the changes, and so on. **As a result, the processed voting data only becomes available several months after the August 31 N-PX submission deadline**, because it takes the service providers a while to process the data, check the quality of their extraction software, and so on. With tagged data, all kinds of summary statistics and data analysis (as well as the actual data behind them) could be automatically available to everyone immediately on September 1, at virtually no cost.

Another objection to the Release questions the feasibility of uniform tagging. I think this concern is unfounded (although I do of course agree that carefully designing an appropriate tagging system is critical for the usefulness of the project). Virtually all votes fall under a few clear categories, so uniform tagging should not be a problem. For example, service providers mentioned above already have classification schemes for various types of votes. It would of course be very useful to also be able to tag companies, directors, and issues in a way that is immediately computer-readable and cross-linkable (linking various identities to CIK identifiers as the Release says, or to the global legal entity identifiers currently under development by a committee of the Financial Stability Board). This may be somewhat more challenging and require more coordination, but (a) I do not think this would be prohibitively hard, except perhaps for a few special cases (e.g., small foreign companies) and (b) even without that additional cross-linking, tagged data would be very useful. Simply requiring the funds to report a company's ticker and CUSIP (as is already done), and slightly refining the way individual votes are reported (e.g., a specific category for "Ratify Auditors"; separate tags for a director's first name, last name, etc. in director elections; and other similar tags for other issues) would go a long way to making the data computer-readable and analyzable.

Yet another objection is that there will be additional costs to turn the structured data into a human-readable form. I think these costs will be trivial: it is much, much easier to go from a structured dataset to an unstructured one than in the opposite direction. Any competent web designer or even a computer science student will be able to write software to translate a structured XML document (or a structured, properly tagged document in another format) to an HTML document (readable with a standard internet browser) or a plain text document (directly human readable) in a matter of hours, or at most days -- and this will only need to be done once (and updated only when the tagging rules are updated for some reason). It will be equally easy to translate these structured documents into formats amenable to analysis, such as Excel spreadsheets. Moreover, I am certain that once structured tagged data becomes available, various web services displaying it and allowing users to interact with it, summarize it, graph it, etc., for free or at a very low cost, will appear very quickly, in part due to the trivially low costs of setting up such services based on structured data.

I have a few additional comments on the Release.

First, the release talks about requiring funds to disclose the number of shares they voted, which they currently do not. I think **this would indeed be very valuable**, in light of such issues as empty voting and over/under-counting of votes, which the SEC is trying to address. For any fund that already has information on how many shares they voted, the cost of adding this number to their disclosure forms will be very low. And any fund that does not have this information will be forced to learn it, which cannot be a bad thing for the integrity of the proxy voting system. Disclosing this information will also be valuable to investors: For instance, if in some closely

contested proxy fight or merger, most funds vote in a particular direction, but it turns out that their numbers of votes are much lower than their holdings (and this difference changes the overall voting outcome), that is something that company and mutual fund shareholders, as well as policymakers, should know about.

Second, in some places **the Release talks about "permitting or requiring" the funds to submit the data in tagged form. I think just "permitting" such disclosures is insufficient**: if some funds report structured data, and some do not, all of the problems I mentioned above will persist. If the rules change in the direction of structured data, it should be made a requirement, not an option.

Finally, the Release proposes requiring the funds to report data "in interactive data format in addition to the traditional format." It would be sufficient to only require them to report it in the interactive format. As I mentioned above, translation from the interactive format to human-readable HTML and plain text formats, as well as to a PDF format for printing, to an Excel format for analysis, etc., can be done automatically, at very low cost – provided the tagging system is set up properly. In fact, should the need arise, I would be happy to write such translation software myself, and make it available to the SEC at no cost (although I suspect the SEC engineers would be able to write this software just as quickly and easily).

Please feel free to contact me at ▉▉▉▉▉▉▉▉▉▉▉▉▉▉ if you have any questions.

Sincerely,

Michael Ostrovsky

cc:

Mary Jo White, Chair
Luis A. Aguilar, Commissioner
Daniel M. Gallagher, Commissioner
Kara M. Stein, Commissioner
Michael S. Piwowar, Commissioner

Norm Champ, Director of Investment Management
Keith F. Higgins, Director of Corporation Finance
Craig Lewis, Director of Economic and Risk Analysis