Act**Blue**

14 Arrow Street
Cambridge, MA 02138

(617) 517-7600 PHONE
(617) 517-7601 FAX

Want Blue States?
www.actblue.com

To: Robert Hickey, Staff Director, Federal Election Commission
From: Matt DeBergalis, Chairman and Founder, ActBlue
Re: Website & Internet Communications Improvement Initiative
Date: 21 July 2009

Dear Sir:

I applaud the Commission's effort to improve the FEC website and disclosure tools.

Much of the Commission's request for public comments revolves around the access to and presentation of disclosure information. This is my particular area of expertise: since 2004, ActBlue has reported over two million individual disclosure events to the FEC. A software engineer by training, I have been variously responsible for architecting ActBlue's novel use of earmarked contributions, designing and implementing the software that processes these individual transactions, and managing the task of integrating that activity into ActBlue's monthly disclosure filings. I have also worked with the equivalent disclosure systems in over 20 individual states, again both from a campaign finance perspective as well as a software engineer.

It is in this context that I submit the following testimony to the Commission. I would welcome the opportunity to expand on these comments by testifying to the Commission at its hearing on this matter next week.

§

I believe the Commission can use a technology refresh as an opportunity to establish the FEC database as the gold standard of Federal compliance data while fostering a thriving ecosystem of independent software tools designed to query the FEC master data for specific information or to conduit specialized analysis. The Commission has a unique ability to serve as a kind of "neutral data warehouse" for compliance information, establishing standard data formats and online protocols, providing canonical identifiers for individuals and vendors that appear in compliance filings, ensuring interoperability, and publishing reference implementations of its own software where appropriate.

A clear data model opens the door to third parties who can build their own independent tools for submitting, analyzing or visualizing committee disclosure data. Some tools may prove useful to many committees or other interested parties; indeed, the Commission should search for ways to encourage developers to publish these tools. Others will be developed privately, perhaps on behalf of a individual committee to answer strategically important questions, or a third-party watchdog searching for hidden trends or archetypes. These tools may potentially remain far from the public's eye.

While the Commission will almost certainly build its own analysis software on top of this new foundation, a rich and well-documented data model relieves the Commission of many of its own software burdens. With appropriate formats and protocols in place that leverage open industry standards and development tools – both free and proprietary –

the Commission is no longer in the unenviable position of gating access to disclosure data behind its own software. In other words, while the Commission may continue to develop and support its own systems for uploading or analyzing disclosure data, these would no longer be required for other groups to develop their own systems.

From this perspective, the data itself reigns supreme and the toolkits are merely supporting players. The Commission is the only custodian of the data itself, and should focus its efforts on providing clear, complete, and well-documented data to its clients.

The Data Warehouse Model

An excellent starting point for defining a clear data model are standard techniques developed for data warehousing in the commercial world. The universe of disclosure data that the Commission intends to publish should be integrated into a comprehensive schema. Such a schema would revolve around "facts" such as contributions to a committee, and "dimensions," which might include committee information, individual donor information, election information, and other metadata relevant to the disclosure toolkit itself.

A successful data warehouse places the focus on individual facts and their supporting dimensions. The commission must target its greatest efforts here. If the data definitions capture the important meanings of each disclosure event, then there are numerous opportunities for a richer analysis framework, either developed in house by the Commission or by independent for-profit and non-profit groups. On the other hand, without a clear model and standards-based access to it, the Commission must carry the burden of developing and maintaining the toolkits used to upload compliance data (such as FECFile) and retrieve information (such as the custom search features of the current website).

In the language of data warehousing, one of the critical roles of the Commission is to canonicalize and aggregate incoming data into standard forms before publishing them in the warehouse. Canonicalization is the process of converting various similar representations into a single, standard value. A good example is converting postal addresses into a standard format with a ZIP+4 code. Once data is canonicalized, the commission may perform one or more aggregate calculations and provide those values to the warehouse alongside the individual disclosure database. In no case should aggregates replace the underlying facts in the warehouse, which are essential for third-party analysis that may be impossible when starting just with the aggregate information. The commission already does much of this canonicalization and aggregation today. The key is to precisely document the rules that are used, so that there is certainty about the final result and an opportunity for third party tools to align their own business logic.

Another critical role for the Commission is to develop standards for coding contributors, employers, industries, locations, purposes, and vendors with unique and consistent identifiers. Continuing with the language of a data warehouse, each of these attributes of a disclosure event is a dimension. Much of the challenge of a warehouse is in

maintaining a comprehensive database of these dimensions, particularly as new values are added over time. This task is essential, though: these dimensions form the backbone of virtually all the sophisticated analyses one might wish to attempt. In contrast to the value of independently-developed front-end tools with distinct strengths and weaknesses, there is no value in competing techniques for identifying repeat contributors or common vendors across multiple disclosures. Indeed, disagreement over whether two donor records refer to the same individual is singularly unhelpful. The Commission's internal system is the natural place for such an effort.

Formats and Protocols

The Commission has a clear opportunity to define standard formats for disclosure data and standard protocols for the transmission of disclosure data. Naturally, these standards should build on modern software engineering best practices. Any new website and disclosure toolkit should accept incoming disclosure data in these formats, and provide canonical disclosure data to the public using them.

Two formatting standards would be particularly valuable: a standard XML representation of campaign finance events, committee records, and communications between the Commission and committees; and an XHTML microformat standard appropriate for use in an interactive website. Microformats are particularly well-suited for committee records and contribution and expenditure data. A successful microformat standard would allow both generic and purpose-built search engines to easily index disclosure data, offering an end-run around many of the search challenges the Commission raises in its RFP.

On the protocol side, it is critical that the Commission replace the closed software systems currently used to upload disclosure data with standard network protocols that third parties can easily tie into their existing systems. This work paves the way for more sophisticated reporting tools, real-time disclosure, and fewer translation errors caused when committees struggle to force disclosure data into FECFile.

Open Systems

Finally, I urge the Commission to publish the disclosure system itself.

The Commission has an opportunity to improve campaign finance disclosure not just at the Federal level, but in states and local municipalities. The improvements I've suggested – and indeed the challenges laid out by the Commission in the RFP – are as applicable to state and local election commissions, many of whom have far fewer resources than does this Commission. While not a formal part of your mission, I would suggest that offering a template for disclosure systems to others in need of these tools is not at all in opposition to the FEC's charter. Standardizing formats and protocols grows the market for tools built against these standards, and will inexorably lead to a greater variety of higher quality tools.

The process should begin with a public collaboration to ensure a robust standard that meets disparate needs. The software components that define the data warehouse schema, accept incoming data, canonicalize and aggregate records, house the disclosure warehouse, support client queries and downloads, and Commission-developed front end tools should all be made available to the public in source form, with clear documentation and change histories. Many of the search and analysis tools currently offered by the Commission, such as the widgets on the current website to view data by House and Senate Elections, can be re-implemented using the new warehouse tools and again be made available in source form for others to modify and build on. And the process by which these tools were developed, including internal deliberations over formats and techniques are all valuable material that also deserve to be made available to the public.

Thank you again for the opportunity to contribute to the Commission's request. I look forward to a productive conversation.

Sincerely,

Matt DeBergalis

Chairman and Founder, ActBlue