

To: rule-comments@sec.gov.  
From: marchywka@hotmail.com  
Subj: comments in response to S7-27-08

Hi,  
I'm responding to the document published at

<http://www.sec.gov/rules/proposed/2008/33-8982.pdf>

33-8982 Nov. 14, 2008 "Roadmap for the Potential Use of Financial Statements Prepared in Accordance with International Financial Reporting Standards by U.S. Issuers" Other Release No.: 34-58960 File No.: S7-27-08 Comments Due: February 19, 2009

with remarks (which are not really IFRS-specific) addressed to the sections "Improvement in the Ability to Use Interactive Data for IFRS Reporting." (III.A.3) and "The Roles of Financial Information" ( III.B.1).

I would request, in order to reach the largest interested audience with the most useful business data, that

1) a documented search/retrieval API for XBRL filings( as well as existing text filings ) be added to the SEC website to provide functions similar to the human interface at

<http://www.sec.gov/idea/searchidea/webusers.htm>

but designed for automated access as described at the NCBI facility for medical literature,

[http://www.ncbi.nlm.nih.gov/entrez/query/static/eutils\\_help.html](http://www.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html)

( sorry for wasting your time if this already exists but it wasn't obvious and I wasn't optimistic for reasons elaborated below. The RFC gives few details of how filings are to be made available at SEC or filer's website. )

and

2)that any accounting rules, IFRS or otherwise, that allow model based valuations, especially those that involve assumptions about forward looking scenarios, also require reporting of model input data sufficient for some independent analysis if not complete replication.

The fullest use of accounting data is made when as much raw data as practical is made available to the most versatile tools available without making unreasonable requests on the user's computer resources. The SEC already provides a database of filings, so most users would have no need to replicate the database locally or try to go to specific company websites to find desired filings. The SEC is moving towards computer readable filings with the XML structure, fixing another problem with current filings for use in automated analysis programs.

However, they do not appear to offer a programmatic search capability or API via which user programs could request and load arbitrary collections of documents. The NCBI eutils interface cited above provides an excellent example of how diverse types of data can be made available to both developers and casual users and it should be adopted by the SEC and other groups for business related data. This approach would service the needs of students, academic researchers, casual investors, and analysts, while allowing developers to repackage the data into "value-added" websites for any audience that the SEC has missed.

Aside from the obvious interest in allowing investors to know what they are buying, I'm basing these suggestions on the notion of the public acting as a sanity check on markets or perhaps "lead generators" for enforcement organizations including but not limited to the SEC. Irregularities or "interesting patterns" can be uncovered by an academic, journalist, or other interested citizen performing ad hoc analysis on statistically meaningful samples of ( large amounts of ) public data. The options backdating problem, for example,

[http://www.thehcmr.org/issue2\\_1/stat\\_scandal.pdf](http://www.thehcmr.org/issue2_1/stat_scandal.pdf)

was discovered using a related approach. I believe the suggestions I'm making facilitate these efforts by a larger group of interested and competent parties either directly or by enabling them to offer reliable value-added websites to others with a minimum of effort.

I have used many of the SEC online access facilities for the US-GAAP text filings that they accept currently. These have been very helpful and the move towards computer readable XML (XBRL) filings will further facilitate automated analysis. The SEC website offers searches by various keys as well as the important full text search and a complete ftp site

<http://www.sec.gov/edgar/searchedgar/ftpusers.htm>

that apparently is free ( except for the real time feeds ) and well organized.

However, I am asking that part of the roadmap for future usage allow access to a wider audience by making filings in the SEC database searchable through an API or interface suitable for programmatic access with non-proprietary tools.

Looking at the current XBRL documentation,

<http://www.sec.gov/spotlight/xbrl.shtml>

it appears to still be oriented towards human interaction and retrieval from an HTML search form. This alone doesn't facilitate all the stated goals,

"Under those proposed rules, financial statement information could be submitted by public companies in interactive data format, and that financial information could then be downloaded directly into spreadsheets, analyzed in a variety of ways using off-the-shelf commercial software, or used within investment models in any of a number of other software formats."

as custom analyses could request many documents at one time selected upon some criteria. The listed resources on this and linked pages seem to be confined to viewers, implying a human audience with no particular programmatic interface.

The "source code" for the viewer is about 33 megabytes of XML files, PHP and Windows DLL's but all that is really needed for automated analysis is a short document describing how to get the filings of interest through a concise query without any HTML formatted for human consumption. It is possible that my suggestions are in fact buried in there somewhere ( presumably the viewer can programmatically retrieve filings from some database ) but it would take a while to determine. A better approach would be to offer and document an interface such as the NCBI database interface,

[http://www.ncbi.nlm.nih.gov/entrez/query/static/eutils\\_help.html](http://www.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html)

which provides a stable tractable means to query and retrieve results in a variety of formats. In particular, the full-text search capabilities are very important for data mining and should be offered as part of this API ( along with the addition of indicies for alphanumeric terms that may be important, as many names contain digits, but are not currently indexed). If my memory is correct, I think in the past I have corresponded with textsearch@sec.gov and this office did not envision any type of API and wasn't sure how integration with IDEA would proceed. I would suggest an integrated search and retrieval API similar to the above NCBI facilities. Rarely would I go to the web site for a specific company to obtain filings any more than I would go to a specific institution to obtain scientific publications when I can just go to the NCBI facilities.

One general comment pertaining to accounting rather than just access methods would be a request that any time a model based valuation is allowed that significant requirements be imposed for reporting the input data sufficient for the reader to at least "sanity-check" the model. Models, aside from being arbitrary and susceptible to bugs or errors, usually make assumptions about the future being similar to the past. The SEC, as opposed to other regulators, has recognized that fortune telling is not an exact science with the now famous "forward looking statements" language and analysts may want to evaluate the role of assumptions about the future in various models. Derived quantities defined by accounting standards, summary numbers like "earnings", are great and precise definition is important for a casual user but with computers it is possible to report details up to the limits imposed by confidentiality issues.

Thank you,  
Mike Marchywka  
Marietta GA

-----  
-----  
Appendix I: Various Supporting Comments, Additional References and  
Details:

=====  
=====

I was debating about not sending this comment until I noted that you changed your HTML format, for example adding the "IDEA" logo,

<http://idea.sec.gov/Archives/edgar/data/70858/000119312509007109/0001193125-09-007109-index.idea.htm>

and I have some scripts to extract information from your human readable HTML that may no longer work. These suggestions would address that potential problem.

My responses to the SEC RFC are a subset of those I elaborated in the case of real estate,

[http://www.federalreserve.gov/SECRS/2008/December/20081210/OP-1338/OP-1338\\_7\\_1.pdf](http://www.federalreserve.gov/SECRS/2008/December/20081210/OP-1338/OP-1338_7_1.pdf)

<http://files.ots.treas.gov/comments/fddad554-1e0b-8562-eb37-34e416089fee.pdf>

or in the case of websites offering technical or scientific data here,

<http://www.mail-archive.com/bbb@bioinformatics.org/msg00143.html>

I am suggesting that all relevant points be considered where applicable to securities filings. The SEC, unlike the real estate regulators, has enabled good use of computer facilities available to much of the interested public. The move to XML filings along with free access to all filings via FTP covers most of my concerns except for the programmatic search capability. The capabilities I'm suggesting have been mentioned in various forums and in particular with regards to the mortgage crisis. One user's comments derived from a private service that no longer exists as a free-access service( note the date ), [http://investorshub.advfn.com/boards/read\\_msg.aspx?message\\_id=13071715](http://investorshub.advfn.com/boards/read_msg.aspx?message_id=13071715) point to how easy it is to find companies with exposure to a given issue using various scripts and a searchable database. I have scripts to perform similar functions by parsing the HTML from the SEC fulltext search but it is cumbersome and there is no assurance the HTML format is stable. A stable API for returning XML or text documents or just a citation list would be very helpful for making the best use of the filings. I've used the SEC full text facilities since they became available and wish to express support for these as far as they go. They were instrumental in finding doomed mortgage companies early on. However, last time I checked, they did not index alphanumeric strings, just pure alpha words.

This makes certain name searches difficult.

The general problem with most information-providing websites seems to be assuming that the immediate consumer for data is a human. It is unlikely, however, that any given web designer can anticipate all the ways an investor or academic researcher would use the data( and of course many commercial sites rely on human interaction with advertising for revenue but this should not be a consideration with the .gov sites). Various programs and scripts could be written to create arbitrary analyses with little effort if a simple API existed.

[ tech digression ... ]

I did try to look for a programmatic way to obtain XBRL filings by downloading some of the code linked from the SEC web pages. For

example, I downloaded some of this code,  
<http://www.sec.gov/spotlight/xbrl/xbrlwebapp.shtml>  
it appears to be a lot of server side code and vendor-specific XML files. If I dig into the files a bit, I find something called docloader with a "config.ini" file that seems to point to this,  
<http://sec.gov/Archives/edgar/xbrlrss.xml>  
to obtain documents but this seems to be, as the name implies, a way to just get the most recent filings and not filings matching arbitrary criteria. If in fact the SEC provides a database API please just post the API somewhere simple ( even a README file in the viewer "sourcecode" would help).  
[ end digression ]

Providing a simple vendor-non-specific API targeted to developers would allow anyone to either do his own analysis or create value-added sites for non- technical audiences and provide facilities that the SEC for whatever reason did not implement. The name "interactive data" does suggest a limited ability to use data via automated methods.

The SEC has taken a lead in highlighting various valuation shortcomings by obsessive use of the term "forward looking" but there are more issues that can be examined if properly organized data is programmatically available. Because many models are forward-looking as the RFC points out ("In this respect, the sensitivity analysis provided under IFRS will be based on forward-looking information.") and anyone modeling the future is free to make the assumptions most favorable to their desired conclusions, it is important that the public have access to input data, not just model output, anytime a model-based valuation is allowed. One humorous model feature is an ability to turn ignorance of the future into an asset by using volatility to add to the value of both put and call options. Clearly, volatility adds no economic value which is the only value that finally matters in aggregate and may be one component of book value that a researcher may wish to investigate. Similar comments would apply to any valuation to which unusual conditions or "qualitative factors" or various "contingencies" or "covenants" or "trigger events" are attached.

Reporting in a decision tree or other codified manner would facilitate better automated analysis. Even better, for financial assets, make sure that CUSIP's are given and that registration documents or prospectuses have encoded all of these issues. Personally, I'd like to see sufficient information to find counterparty loops and similar issues. Counterparty identification for any obligation, liability, or financial asset is a great help in "netting out" the economic component, if any, on a balance sheet.

To summarize, I look forward to the day when all legal documents are computer readable and formatted similarly to the SEC XBRL documents. Now is the time to make some of these documents available in a flexible programmatic manner from a centralized data source so the machine-readable aspect can be exploited and evaluated.

Thanks again if you read this far :)

Mike Marchywka