



Wharton Research Data Services

The Wharton School
University of Pennsylvania

3819 Chestnut Street
Suite 300
Philadelphia, PA 19104.6301

Tel: +1.215.746.4331

Fax: +1.215.573.6073

May 26, 2020
Ms. Vanessa Countryman
Secretary
U.S. Securities and Exchange Commission
100 F Street NE
Washington D.C. 20549

Re: **Notice of Proposed Rule on Market Data Infrastructure, Securities and Exchange Act Release No. 88216 (February 14, 2020) (File No. S7-03-20)**

Dear Ms. Countryman,

Wharton Research Data Services (“WRDS”) respectfully submits this comment letter on behalf of WRDS, and the academic community, in response to the Proposed Amendment of request for comment, that would amend Rules 600 and 603 and adopt Rule 614 of Reg. NMS.

Executive Summary

On February 14, 2020, the Securities and Exchange Commission (“Commission”, “SEC”) published Proposed Rule 34-88216 detailing amendments to market data infrastructure, consolidated market data and its distribution. In Proposed Rule 34-88216, the Commission solicited comment from the academic community concerning the social welfare costs of the discontinuation of NYSE’s TAQ product¹. WRDS, as the sole distributor of TAQ to the academic community, and as a leading member of the academic community, is uniquely positioned to comment on these issues.

The distribution and access of trade and quote data is essential for the study of many important topics in the academic literature. These topics, including price efficiency, liquidity, short selling, high-frequency trading, market fragmentation, et al, are essential for providing a more complete understanding of the welfare, function, and efficiency of financial markets. TAQ data is currently the primary source of widespread trade and quote data to the academic community, with 107 universities and government agencies subscribing through WRDS, including the Commission itself.

¹ The Commission’s solicitation for comment #244 in Proposed Rule 34-88216 was as follows:

Would the proposed amendments result in the interruption of data available for research by the academic community and investors, such as TAQ data? If so, the Commission requests that commenters provide relevant data and analysis to assist us in determining the incremental social welfare cost of such interruption of data to the academic community and investors.

We welcome the new consolidated competitors model, which more accurately reflects the reality of financial markets. This is a needed next step in reflecting the inaccuracies of the NBBO as currently imagined. However, with this new model, TAQ data, the primary source of top of book U.S. equity order-level data, will disappear.

Because of this reality, we have come to believe the following:

- NYSE’s TAQ product is licensed to the academic community at a steep discount to its true cost - despite this discount, many universities are unable to afford TAQ data
- The end of TAQ top of book data would cause commensurately large costs to both the academic community and aggregate social welfare
 - Discontinuation of U.S. equity top of book data would impact
 - The advancement of understanding and contributions provided by approximately 6% of all published academic work in the field of finance.
 - The development of policy and regulations underpinned by empirically supported scholarly research. It should be noted that, the SEC’s Proposed Rule itself draws its preliminary beliefs from knowledge accrued in 37 different academic studies.

With these facts in mind, we propose the following:

- Consolidated market data, as newly defined by the Commission, be made free for academic use.
- Allow Competing Consolidators to provide consolidated market data to a registered “Academic Aggregator” (e.g. WRDS) who will be willing to support the low-margin academic research community at the high level necessary for quality research.

Background

WRDS welcomes the new competing consolidator model. Preliminary academic work² has begun to acknowledge the ineffectiveness of Securities Information Processor (“SIP”) generated data for academic research. Geographic latency greatly impacts common metrics used by academics³, illustrating a market reality where there may be up to 13 different interpretations of the NBBO at a time⁴. This disparity greatly impacts academic research, with computational costs for geographic latency adjustments to create each exchange’s relative NBBO (“RBBO”) record taking 9.6 hours per day of data⁵. The consolidated competitor model, by accepting these facts, embraces a new and more accurate system for distributing market information.

Wharton Research Data Services

Wharton Research Data Services (WRDS) has served the academic research community since 1997 with its data research platform. Today, 480 leading academic institutions located in 35 countries rely on WRDS to access finance and accounting data that underpin a significant portion of empirical academic

² Holden, Pierson, and Wu (2020), Hasbrouk (2019), Bartlett and McCrary (2019).

³ Holden, Pierson, and Wu (2020) find usage of SIP data compared to estimates of direct feeds impacts estimation of many commonly used liquidity, HFT, and other measures.

⁴ Holden, Pierson, and Wu (2020) find estimates of alternate NBBOs adjusting for geographic latency differ from SIP published NBBO quotes by 22% per day.

⁵ Over the course of the history of participant timestamps in TAQ data, geographic latency adjustments to create each exchange’s RBBO would take a WRDS user ~10,800 hours (or ~1.25 years) of computing time.

research. In 1998, WRDS began hosting NYSE Trade and Quote (TAQ) data as one of the 500+ datasets published by over 50 data vendors and hosted by WRDS. As of January 2020, approximately 22% of WRDS academic subscribers license TAQ data directly from NYSE and access the data via the WRDS platform. TAQ is a computationally intensive data set that requires processing on dedicated servers by WRDS.

WRDS hosts TAQ in order to provide a source of trade and quote data to the academic community. Many institutions may find it difficult to store, process, and manage such a historically large data product on their own internal servers, and duplicating these efforts for hundreds of universities would be even more costly. WRDS offers a centralized service providing a single dedicated, managed, and supported location to access and analyze TAQ data. Doing so provides substantial cost savings to not just higher tier research institutions, but any institution actively involved in academic research, leveling some of the resource inequalities in higher education. Leveling these inequalities further allows a greater number of researchers to access the data, generating more insight and knowledge than otherwise might be created.

WRDS also provides user support for its hosted data, answering over 560 questions by academic users related to TAQ data in 2019. A team of economics and finance PhDs work to support the data WRDS provides, addressing issues ranging from technical concerns, to data generation, to research methods. TAQ support queries range from the technical to the academic—how to most efficiently compute statistical analyses on TAQ data to the historical start date of odd lot trade submissions to the SIP.

Our singular position in the academic sphere, as well as the services we provide, makes WRDS uniquely qualified to understand how TAQ is used within the academic community, and what social welfare costs a loss of such data would generate.

The Importance of TAQ to the Academic Community

The body of research literature that has been developed as a result of access to TAQ data by the academic community is highly significant. There are two categories of research papers to consider when measuring the volume of research tied to TAQ data. The first category to explore is the collection of research considered as “working papers” where the research is still in development and not yet published in a journal. One of the most notable research repositories for such a collection can be found at [ssrn.com](https://www.ssrn.com). SSRN was acquired by Elsevier in 2016 and is widely recognized for its multi-disciplinary collection of working papers. Scholarly researchers from academia, government/non-profits and industry all post working papers to SSRN in order to create awareness and to gather feedback on early stage research. SSRN is an open-access platform and provides full-text search capabilities. By searching on “TAQ” across all papers hosted by SSRN, one can retrieve 1,228 working papers⁶ that cite TAQ in their research. Included in this result are papers that cite “TAQ” where the authors may not have used “TAQ data” so the results may be slightly inflated.

The second category to consider for research enabled by TAQ data are the published papers found in academic journals. As the body of research published in finance, accounting and other business-related journals is vast⁷, we have created a subset of journals considered “top tier” journals that are highly ranked by the Financial Time’s Research Ranking and focused on only finance and accounting. Among these journals, TAQ data is used in 556 papers published from 2000 to 2018, consisting of 6% of total published academic research. The topics these papers cover are numerous, including liquidity, high-frequency trading, short selling, market quality, risk and return, equity issuance, market efficiency, and information acquisition.

⁶ Accessed February 27, 2020

⁷ See, for example, <https://www2.cabells.com/>

Finally, the amount of contact WRDS users have with TAQ data illustrates its importance. Over 100 academic institutions subscribe to some form of TAQ data through WRDS. Researchers affiliated with these institutions used WRDS' servers and website to access TAQ data over 575,000 times in 2019. Further, in 2019, researchers ran analyses on WRDS servers using approximately 3.4 petabytes of TAQ data.

Costs to the Academic Community

NYSE provides academic access to their TAQ product at a steep discount. The average cost per academic institution is approximately a 90% cost discount compared to industry participants.⁸ In addition to this pricing discount, WRDS charges a flat platform fee for TAQ subscribers, unlike most cloud storage providers, to incentivize unlimited usage and provide cost certainty for academics at much greater expense to WRDS' computing resources. Under the current consolidated market data ecosystem, academic users can access data for a fraction of its market price. NYSE has been extremely open-minded in their support of academic use with very little compensation in return. The use case is enduring knowledge.

Yet despite this deep discount in pricing from NYSE and storage cost subsidies provided by WRDS, many academic institutions find TAQ data expensive. Of the current eligible subscribers to TAQ data through WRDS, 94% are R1 or R2 classified institutions by the Carnegie Classification System¹⁰. The dearth of less research-intensive institutions (i.e. R3, M1, M2, M3) indicates that schools below the highest tiers of research funding find TAQ too costly.

A top of book replacement product for TAQ, despite possible cheaper costs from competing consolidators¹¹, even if it did not consist of more consolidated market data¹², would incur significantly greater storage costs than TAQ due to changes in the definition of "round lot". From the Commission's own estimates¹³, combined with the average closing price in the prior day for all symbols traded on NYSE ARCA and NYSE, we anticipate an increase in the storage costs of TAQ of 700%¹⁴. An increase in storage costs of this magnitude, regardless of changes to current pricing for a top of book data product, would increase costs to the academic community or demand greater subsidies from WRDS and, assuming they become a competing consolidator, NYSE. These costs would bear their own welfare loss, with universities unsubscribing from these data with the rise in storage costs.

WRDS recognizes that TAQ data is critical to academic community's understanding of financial markets. In these historic times of economic and social instability, the study of financial markets is even more important. Yet an interruption in the availability of TAQ to the academic community would cause a great loss of analysis. This loss of will adversely impact not just academic research but progress in education and insights in policy—insights which the Commission currently draws on. WRDS currently provides

⁸ Based on historical pricing information provided by NYSE. WRDS understands this number to still be approximately correct.

¹⁰ The Carnegie Classification of Institutions of Higher Education (n.d.). About Carnegie Classification. Retrieved from <http://carnegieclassifications.iu.edu/>.

¹¹ The Commission believes this may be true, though there is the possibility that it is not. See Section VI.C.2(b) of Release No. 34-88216; File No. S7-03-20.

¹² The Commission admits this may occur on page 463 of Release No. 34-88216; File No. S7-03-20

¹³ Table 3 of Release No. 34-88216; File No. S7-03-20

¹⁴ Based on analysis of NYSE and NYSE ARCA sample order book data compared to the size of the same day's TAQ NBBO Eligible Quotes file. Sample data accessed from <https://www.nyse.com/market-data/real-time/integrated-feed>

these resources for the academic community and WRDS will always seek to do whatever it can to help researchers gain the empirical data and tools they need to answer important questions. But without assistance from the Commission, we believe the future of this highly important segment of academic research to be at risk.

Proposals

As established above, the continuation of TAQ data is essential to sustain impactful academic research. WRDS is uniquely positioned to maintain, support, and disseminate TAQ to the academic community. If the Commission's Proposed Order comes into effect as currently written, it is the opinion of WRDS that the academic community will be adversely impacted. WRDS proffers two solutions that can each resolve this issue.

Consolidated Market Data Free for Academic Use

The costs of studying of financial markets are currently born by major universities, government agencies, and researchers themselves, while the financial benefits are not reaped by these institutions and academics but by society at large. The foundations of economics call for subsidizing positive externalities like this, yet the costs of data acquisition, storage, and analysis are only rising. The SEC could, and should, defray these costs by making access to consolidated market data free for academic use. By doing so, the Commission would ensure that research on public equity markets is preserved both with respect to this proposed rule, as well as any other future changes in financial market regulation or structure. WRDS believes that making consolidated market data free for academic use would not only ameliorate any increase in costs for TAQ data in the future but allow access to institutions with lower tiers of research funding.

An "Academic Aggregator" to Support Academic and Government Research

Competing Consolidators, as defined by the Commission, may have little incentive to support academic subscribers, who offer low potential margins combined with high support needs. WRDS proposes the Commission create an additional class of "Academic Aggregator", who can act as an intermediary for the academic community and the Consolidators; assuring continuity of data availability and research-level support to assure ongoing consistent access to top of book data to the research and oversight community. WRDS would be interested in acting as an "Academic Aggregator", providing its current level of support for NYSE's TAQ, to Consolidated Market Data while distributing it at a lower cost to the academic community.

Summary

The social welfare costs of an interruption in TAQ are likely large—6% of papers in top peer-reviewed journals cite NYSE's TAQ data, while the Commission itself references 37 of them. Losing the source of future knowledge will not be healthy for financial markets. WRDS has played an important role in the academic community by a research computing platform allowing over 100 schools access to TAQ data. Consequently, WRDS is uniquely positioned to observe the costs of providing Consolidated Market Data to the academic community and to defray some of them, should more occur. Despite this position, there is no guarantee that TAQ, another top of book data product, or any Consolidated Market Data will be priced low enough to continue uninterrupted access to the entire academic community. The concept of an "Academic Aggregator" is born from this reality—for research to flourish using newly designated Consolidated Market Data, shared resources focused on advancing research must be made available to the

academic community either through designation by the Commission or discounted academic use. We feel it is unrealistic to expect Competing Consolidators to price their service at a level acceptable to the academic community while providing the level of support necessary for academic work. An “Academic Aggregator” or any other host of free or discounted data would be able to provide these services reasonably tied to their costs, ensuring the proliferation of academic research and ameliorating the social cost of an interruption in TAQ data.

We greatly appreciate the Commission’s consideration of these proposals.

Signed,

Robert W. Holthausen
Professor of Accounting and Finance
Nomura Securities Company Professor
Editor, *Journal of Accounting and Economics*
The Wharton School
University of Pennsylvania

Robert Zarazowski
Managing Director
Wharton Research Data Services
The Wharton School
University of Pennsylvania