

U.S. COMMODITY FUTURES TRADING COMMISSION
 Office of the Inspector General
 Office of Audits



**REVIEW OF CFTC'S DATA GOVERNANCE PROGRAM:
 INTEGRATED SURVEILLANCE SYSTEM**

**REPORT NUMBER: 18-AU-07
 May 7, 2019**

TABLE OF CONTENTS

EXECUTIVE SUMMARY 2

ABBREVIATIONS 7

APPENDIX A 8
DATA GOVERNANCE FRAMEWORK DEVELOPMENT AND STATUS

APPENDIX B 12
ISS DATA REQUIREMENTS AND STAKEHOLDER VALUE

APPENDIX C 20
ETL PROCESS FOR DATA STAGING

APPENDIX D 24
CHANGE MANAGEMENT POLICY AND PROCEDURES

APPENDIX E 25
SECURITY AND [REDACTED]

APPENDIX F 28
SECURITY CONCERN FOR SIMILAR LEGACY APPLICATIONS

APPENDIX G 30
BACKGROUND, OBJECTIVE, SCOPE AND METHODOLOGY

APPENDIX H 34
ISS BACKGROUND AND HISTORY

APPENDIX I 37
NOTICE OF FINDINGS AND RECOMMENDATIONS

APPENDIX J 38
MANAGEMENT’S COMMENTS



U.S. Commodity Futures Trading Commission
Office of the Inspector General
Three Lafayette Centre
1155 21st Street, NW, Washington, DC 20581

TO: J. Christopher Giancarlo, Chairman
Brian D. Quintenz, Commissioner
Rostin Behnam, Commissioner
Dawn Stump, Commissioner
Dan Berkovitz, Commissioner

FROM: Miguel A. Castillo, *CPA, CRMA*
Assistant Inspector General for Audits

DATE: May 7, 2019

SUBJECT: Review of CFTC's Data Governance Program
Information Surveillance System (ISS)

Executive Summary

Why We Conducted the Audit

The Office of the Inspector General (OIG) reviewed CFTC's Data Governance¹ program maturity and selected practices pertaining to the Integrated Surveillance System (ISS). This system, developed in the late 1990s, hosts confidential and sensitive market data collected pursuant to CFTC regulations.² ISS supports CFTC market surveillance, market research, public reports (including the CFTC Commitment of Traders (COT) Reports), and other CFTC mission critical activities. See [Appendix H](#) for a timeline description and history of ISS.

The objective of our audit was to assess the maturity of CFTC's Data Governance program and corresponding practices as applied to facilitate the maintenance of ISS. Specifically, we evaluated ISS Data Governance program practices for (1) defining business requirements, (2) extracting, transferring, and loading data (ETL), (3) managing changes, (4) maintaining stakeholder value, and (5) securing data. [Appendix G](#) provides details of our general audit background, objective, scope, and methodology.

¹ Data Governance is the specification of decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, storage, use, archiving and deletion of information. It includes the processes, roles, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals.

² 17 CFR Parts 16 and 17. We did not evaluate whether CFTC Part 16 and 17 regulations could be improved. The OIG may consider this matter in the future.

What We Found

We performed a SWOT (strengths, weaknesses, opportunities, and threats) analysis to assess CFTC's Data Governance program. As presented in Illustration 1, CFTC's program exhibited a low maturity level, displaying numerous weaknesses, opportunities, and threats. We note CFTC is currently working towards an improved Data Governance framework with benchmark attributes. See [Appendix A](#) for further detail.

Since ISS represents a baseline for improvement to the CFTC's Data Governance program, our analysis, as detailed below, provides insight for selected data management practices:

Strengths: The ISS database team followed CFTC Office of Data and Technology (ODT) policies and procedures for ISS change management processes, including obtaining required approvals and conducting security impact assessments. As noted in our prior FISMA review, this capability is a consistent strength for ODT management and no issues came to our attention. See [Appendix D](#) for further detail.

Illustration 1: SWOT Analysis of Data Governance Program Capabilities.



Weaknesses: ISS data is considered a valuable resource across CFTC mission divisions and offices,³ but shows declining usefulness to CFTC operations. See [Appendix B](#) for details of our analysis. Additionally, the collection and maintenance procedures for ISS data are resource intensive and subject to errors. Thus, CFTC may need to consider the current cost, effectiveness, and reliability of ISS data cleansing as used internally, and as the basis for CFTC's external market reports. See [Appendix C](#) for more details.

Opportunities: Our analysis in its entirety supports either updating the existing ISS platform or, following a consideration of costs and benefits, migrating to an updated platform to enhance operational efficiencies while minimizing security threats.

Threats: Our analysis shows that the ISS database application does not comply with federal [REDACTED] requirements for securing federal systems. We found this issue especially concerning given the risk of exfiltration of confidential market and privacy information. We note that CFTC maintains other legacy applications on the same archaic platform, and this may pose similar [REDACTED] risks. See [Appendix E](#) and [Appendix F](#) for our detailed analyses.

We conclude that ISS is less useful today than it was twenty years ago. Given that ISS was developed in the late 1990s, and that CFTC's markets have grown exponentially since that time, some degree of obsolescence may be expected. However, we believe adherence to an effective Data Governance program throughout its lifespan would have guarded against ISS obsolescence impacting CFTC operations, as well as the security [REDACTED] concerns we noted. Given the issues currently existing with ISS, any inability to update and adapt ISS with regard to substance, format, usability, and appropriate data security safeguards (all through appropriate Data Governance processes) may support a decision to migrate to a more modern and efficient technical solution. We realize a thorough analysis of the associated costs and benefits for each option will be necessary.

³ ISS is used by the CFTC Division of Clearing and Risk, the Division of Market Oversight, the Division of Enforcement, the Office of Chief Economist, and the Division of Swap Dealer and Intermediary Oversight.

What We Recommend

We recommend that CFTC:

1. Set a timeframe to fully implement plans for its Data Governance framework and, if not already, synchronize with the goals outlined in the Federal Data Strategy and Open Data Government Act requirements;
2. Update business requirements for ISS and incorporate stakeholder expectations in future ISS versions as a part of an [Enterprise Architecture](#) that aligns with mission operations;⁴
3. Modernize the ISS to enhance the traceability, efficiency, and error handling of ETL processes, which will require a determination whether to update the ISS platform to achieve these goals or, based on a consideration of costs and benefits, to migrate to an updated platform;
4. Given current federal security standards, re-evaluate [REDACTED] requirements for ISS; and
5. Review security risks of other legacy applications and assure compliance with federal information security standards.

How Management Responded

CFTC conveyed its commitment to address all recommendations with appropriate stakeholders. Specifically, CFTC plans to:

- Formulate a data governance framework;
- Address stakeholder business requirements for ISS;
- Upgrade data transmission standards and enhance ETL practices; and
- Ensure security compliance for legacy systems such as ISS.

⁴ In December 2018, we published an audit report addressing issues with [CFTC's Enterprise Architecture program](#).

Management's planned actions are responsive to the recommendations. Management's concurrence and detailed response are presented in [Appendix I](#) and [Appendix J](#), respectively. If you have any further questions, please contact me at (202) 418-5084, or Branco Garcia, lead auditor, at (202) 418-5013.

Cc: Michael Gill, Chief of Staff
Kevin S. Webb, Chief of Staff
John Dunfee, Chief of Staff
Daniel J. Bucsa, Chief of Staff
Erik F. Remmler, Chief of Staff
Anthony C. Thompson, Executive Director
Daniel Davis, General Counsel
John L. Rogers, Chief Information Officer
Srinivas Bangarbale, Chief Data Officer
Naeem Musa, Chief Information Security Officer
Melissa Jurgens, Acting Chief Privacy Officer
A. Roy Lavik, Inspector General
Judith A. Ringle, Deputy Inspector General and Chief Counsel

Abbreviations

CCB	Change Control Board
CI	Configuration Items
CME	Chicago Mercantile Exchange
COT	Commitment of Traders
DCMs	Designated Contract Markets
DMB	Data Management Branch
ETL	Extract, Transform, and Load
FCMs	Futures Commission Merchants
FILAC	Filings and Actio3wwns
FIXML	Financial Information Exchange Markup Language
GAO	US Government Accountability Office
ICE	Intercontinental Exchange
ISACA	Information Systems Audit and Control Association
ISS	Integrated Surveillance System
IT	Information Technology
ODT	Office of Data and Technology
OIG	The Office of the Inspector General
OMB	Office of Management and Budget
SWOT	Strengths, Weaknesses, Opportunities, and Threats

Appendix A

Data Governance Framework Development and Status

In March of 2018, the President set “Leveraging Data as a Strategic Asset” as a “Cross-Agency Priority goal.”⁵ The Federal Data Strategy consists of principles, practices, and action steps to deliver a consistent and strategic approach to federal data stewardship, access, and use.⁶ The principles are considered a timeless and enduring framework for agencies. While the practices are actionable, action steps will be strategically selected for agencies to implement in any given year.

Illustration 2: Federal Data Strategy Goals.

The four Federal Data Strategy areas for exploration

- Enterprise Data Governance**
Set priorities for managing government data as a strategic asset, including establishing data policies, specifying roles and responsibilities for data privacy, security, and confidentiality protection, and monitoring compliance with standards and policies throughout the information lifecycle.
- Access, Use, and Augmentation**
Develop policies and procedures that enable stakeholders to effectively and efficiently access and use data assets by: (1) making data available more quickly and in more useful formats; (2) maximizing the amount of non-sensitive data shared with the public; (3) leveraging new technologies and best practices to increase access to sensitive or restricted data while protecting privacy, security, and confidentiality, as well as the interests of data providers.
- Decision Making & Accountability**
Improve the use of data assets for decision-making and accountability for the Federal Government, including both internal and external uses. This includes: (1) providing high quality and timely information to inform evidence-based decision-making and learning; (2) facilitating external research on the effectiveness of government programs and policies which will inform future policymaking; and (3) fostering public accountability and transparency by providing accurate and timely spending information, performance metrics, and other administrative data.
- Commercialization, Innovation, and Public Use**
Facilitate the use of Federal Government data assets by external stakeholders at the forefront of making government data accessible and useful through commercial ventures, innovation, or for other public uses. This includes use by the private sector and scientific and research communities, by state and local governments for public policy purposes, for education, and in enabling civic engagement. Enabling external users to access and use government data for commercial and other public purposes spurs innovative technological solutions and fills gaps in government capacity and knowledge. Supporting the production and dissemination of comprehensive, accurate, and objective statistics on the state of the nation helps businesses and markets operate more efficiently.

As a primary challenge, the Federal Data Strategy cites robust, integrated approaches to using data to deliver on mission, serve customers, and steward

⁵ <https://www.whitehouse.gov/wp-content/uploads/2018/03/Presidents-Management-Agenda.pdf>.

⁶ <https://strategy.data.gov/>.

resources while respecting privacy and confidentiality. As an opportunity recognized, enterprise-wide Data Governance strategies have the ability to enable government data to be accessible and useful for the American public, businesses, and researchers; and to improve the use of data for decision-making and accountability for the Federal Government, including for policy-making, innovation, oversight, and learning.

In January 2019, the Foundations for Evidence-Based Policymaking Act⁷ called for CFO Act agencies to submit annually to the Office of Management and Budget (OMB) and Congress a plan for identifying and addressing policy questions relevant to agency-specific programs, policies, and regulations. The plan must include: (1) a list of policy-relevant questions for developing evidence to support policymaking, and (2) a list of data for facilitating the use of evidence in policymaking, among other requirements. While CFTC is not a CFO Act agency, it may wish to develop a similar plan as a best practice. Included within the Foundations for Evidence-Based Policymaking Act is the Open Government Data Act, which requires agencies (including CFTC) to: (1) develop and maintain a comprehensive data inventory (metadata⁸) for all data assets created by or collected by the agency, and (2) designate a Chief Data Officer who shall be responsible for lifecycle data management and other specified functions.⁹

Data Governance Key Practices Compared to CFTC Practices

Although a government-wide federal Data Governance framework is in progress, the U.S. Government Accountability Office previously identified common key practices for establishing effective Data Governance structures.¹⁰ Using private organizations and government governance models,¹¹ GAO noted many of these models promote a

⁷ Public Law No: 115-435, Foundations for Evidence-Based Policymaking Act of 2018, Title II - Open Government Data Act (Jan. 14, 2019).

⁸ The term 'metadata' means structural or descriptive information about data such as content, format, source, rights, accuracy, provenance, frequency, periodicity, granularity, publisher or responsible party, contact information, method of collection, and other descriptions.

⁹ PL No. 115-435, sec. 202, amends 44 U.S.C. § 3504(b) to "make data open by default" and amends 44 U.S.C. § 3520 to require the designation of Chief Data Officers. Both provisions apply to CFTC by application of 44 U.S.C. § 3502(1).

¹⁰ US GAO, Data Act: OMB and Treasury Have Issued Additional Guidance and Have Improved Pilot Design but Implementation Challenges Remain, GAO-17-156 (Dec 8, 2016).

¹¹ Id. GAO used organizations that endorsed establishing and using a governance structure to oversee how data standards, digital content, and other data assets are developed, managed and implemented. Based on these selection factors, they drew on work from the following organizations to help them identify data governance key practices: American Institute of Certified Public Accountants, American National Standards Institute, Carnegie-Mellon University-Software Engineering Institute, Data Governance Institute, Data Management Association International, Oracle, National Association

common set of key practices that include establishing clear policies and procedures for developing, managing, and enforcing data standards. A common set of key practices, endorsed by standards setting organizations, recommend that Data Governance structures should include the key practices shown below. We compared these key practices with practices in ODT’s Master Data Governance document and applicable CFTC procedures and noted distinct differences in Data Governance enterprise framework approaches (see Illustration 3).

Illustration 3: Comparative Data Governance Attributes Between Federal Government Data Maturity Model and CFTC ODT Management.

Key Practices for Data Governance Structures	Current CFTC Governance Policy/Procedures
<ol style="list-style-type: none"> 1. Developing and approving data standards. 2. Managing, controlling, monitoring, and enforcing consistent application of data standards. 3. Making decisions about changes to existing data standards and resolving conflicts related to the application of data standards. 4. Obtaining input from stakeholders and involving them in key decisions, as appropriate. 5. Delineating roles and responsibilities for decision-making and accountability, including roles and responsibilities for stakeholder input on key decisions. 	<p><u>Master Data Governance</u> states governance [evidenced] in Microsoft Master Data Services and is the responsibility of the Data Management Branch (DMB). DMB governs which entities are considered master data and review the data stewardship of the master data. DMB coordinates with the entity data stewards to ensure the data itself is maintained. DMB will also review applications for new entities and for alterations to the schemata for existing entities. DMB acts as the gatekeepers to MDS and reviews applications to create new entities.</p> <p><u>CFTC Procedure: Evaluate Information Governance Questionnaires for Proposed New Systems, System Changes or Data Collections</u> provides for intra-CFTC collaboration for new data collections. Its focus is the privacy and security of data.</p> <p>As a matter of practice, CFTC’s DMB leads, develops, and implements guidebooks, and validation rules for its internal governance. Additionally, DMB seeks stakeholder/industry feedback for setting data standards.</p>

of State Chief Information Officers, National Institute of Standards and Technology, Digital Services Advisory Group and the Department of Education-Privacy Technical Assistance Center.

ODT is working towards a Data Governance enterprise framework that includes each of the key GAO cited practices described above (see Illustration 4).

Illustration 4: CFTC ODT Data Framework.



According to ODT, some parts are functional, such as the Data Steering Committee and the Data Officers Technical Working Group. ODT also is working with the Chairman’s office to lay out a renewed data plan for the agency as shown in Illustration 4. Given ongoing resource constraints, ODT anticipates the release of data-specific policies to follow.

Recommendation

1. We recommend that CFTC set a timeframe to fully implement plans for its Data Governance framework and, if not already, synchronize with goals outlined in the Federal Data Strategy and Open Data Government Act requirements.

We believe that fully implementing a recognized framework as shown in Illustration 4 would also enhance the efficiency and effectiveness of back-end operations to deliver value to ISS stakeholders. The sections that follow provide insight into current operations for managing ISS data.

Appendix B

ISS Data Requirements and Stakeholder Value

As previously highlighted, GAO's Key Practices on Data Governance structures stresses obtaining input from stakeholders and involving them in key decisions as a key Data Governance practice.¹² Additionally, information governance and management knowledge organizations, such as the Information Systems Audit and Control Association,¹³ place emphasis on delivering value to stakeholders from an enterprise perspective.

The ODT does not maintain a policy for re-evaluating historical requirements for legacy systems. In order to gain an understanding whether ISS meets current business requirements and offers data accuracy and use-ability, and in order to report user value, we researched public comments on CFTC regulations and outside inquiries to CFTC regarding ISS; conducted an internal ISS user survey; and evaluated Google analytics for published CFTC reports that use ISS data.

Requirements and Comments

We note that CFTC regulations (Part 17) require large trader reporting¹⁴ to be in a single file with each record in an 80 character format.¹⁵ For Part 16 data, CFTC requires a FIXML format;¹⁶ a widely adopted format for derivatives post trade

¹² See fn.9.

¹³ COBIT framework, *Enabling Information*, www.ISACA.org

¹⁴ <https://www.cftc.gov/IndustryOversight/MarketSurveillance/LargeTraderReportingProgram/ltrp.html>
<https://www.cftc.gov/IndustryOversight/MarketSurveillance/LargeTraderReportingProgram/ltrformat.html>
<https://www.cftc.gov/LawRegulation/DoddFrankAct/Rulemakings/XXXII.LargeSwapsTraderReporting/index.htm>

¹⁵ In the beginning of computer technology, a line of an IBM punched card could consist of only 80 characters. The widespread computer terminals such as IBM 3270 followed this limitation, their monitors could show only 80 characters per line (CPL) (but with the various number of lines), though with some terminals this number was either reduced by half to 40 CPL, limited to 64 CPL (SWTP CT-64, with 16 lines), or optionally increased to 132 CPL (DEC VT100 family, with 14 lines). Such line lengths have been carried over into text modes of personal computers.

¹⁶ FIXML (Financial Information Exchange Mark-up Language) is the XML encoding used within FIX. FIXML is widely adopted for derivatives post trade clearing and settlement globally. FIXML is also used for reporting.

clearing and settlement globally. A search of public comments¹⁷ relating to “data” revealed intensity towards the use and availability of data. While we did not find comments related to “Part 16 and 17 data” specifically, we note relevant excerpts from 2 comments associated with “large trader data”. They are as follows:

- ...we have 145 employees, including dedicated developers for regulatory systems, who work on Market Regulation duties. We have 40 staffers dedicated to market surveillance. The annual direct cost of maintaining this self-regulatory program is over \$30 million, with an additional \$5-7 million in regulatory technology support as well as indirect support from other departments...
- Data has not kept pace in reporting on the profound changes in market...supplementation and refinement of Commission data are necessary...Reports is produced once a week and provides only a very summary view of the market...the COT Report data has additional inherent limitations...primarily classification scheme in use.

¹⁷ <https://comments.cftc.gov/PublicComments/CommentList.aspx>

Google Analytics for CFTC External Reports

Static reports display data that is relevant to a specific point in time, are shared easily, and provide consistent information for a defined period, creating a unified perspective. Time-based reports such as CFTC’s weekly and monthly reports generated from Part 16 and 17 submissions are static. We studied Google analytics statistics of CFTC’s reports that contained part 16 and 17 data. Between October 1, 2015, and June 30, 2018 (1003 days), CFTC.gov market reports containing part 16 and 17 data were viewed both domestically and internationally. COT (Large Trader) report viewership averaged .3 daily with a maximum average of 31. Cotton on Call reports were viewed on average .11 times daily with a maximum average of .56. Bank participation reports were viewed the least time averaging .09 views daily with a maximum average of .22 views.

As presented in illustration 5, we noted that viewers in the United States were more so interested in the COT reports. There was much less interest in Cotton on Call reports and Bank Participation reports were viewed primarily in New York.

Illustration 6 shows interest in CFTC reports by other countries. International viewers were more interested in the COT reports. While

Illustration 5: Google Analytics CFTC Reports: Summary Views within the US Domestic Boundaries.

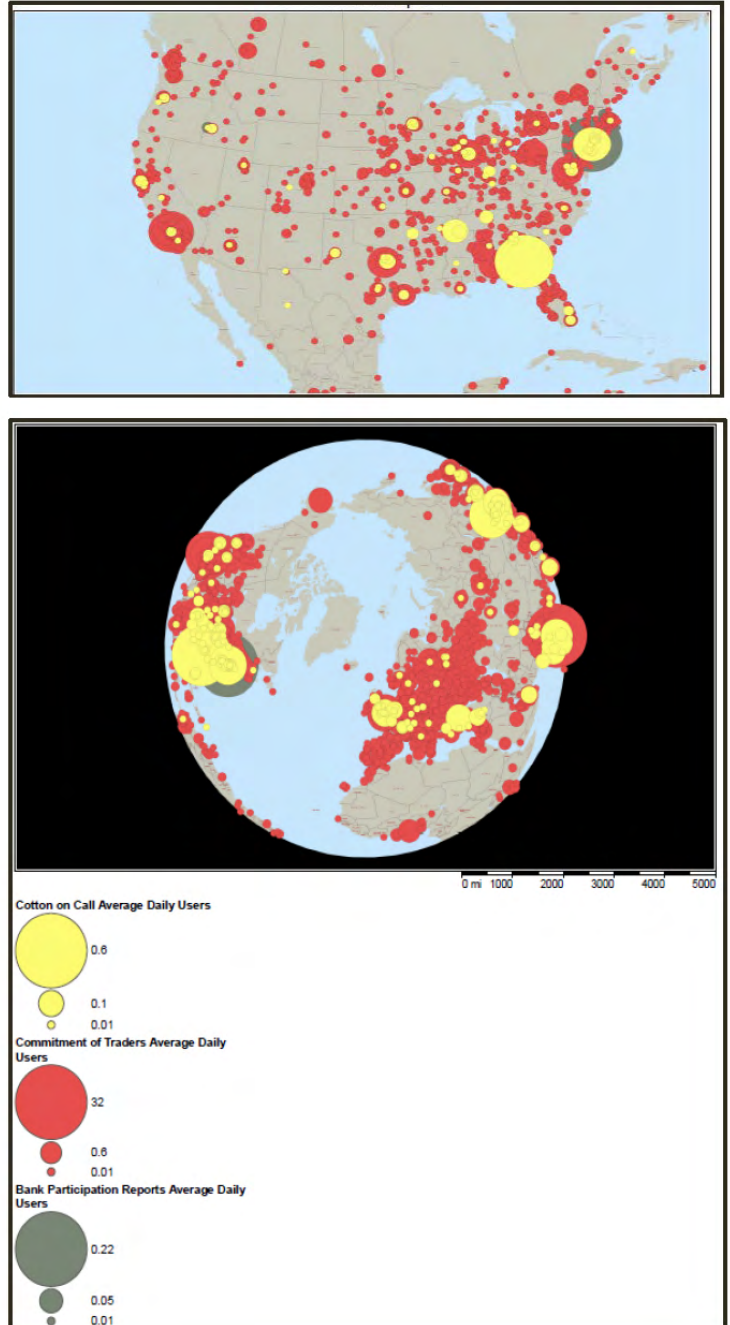
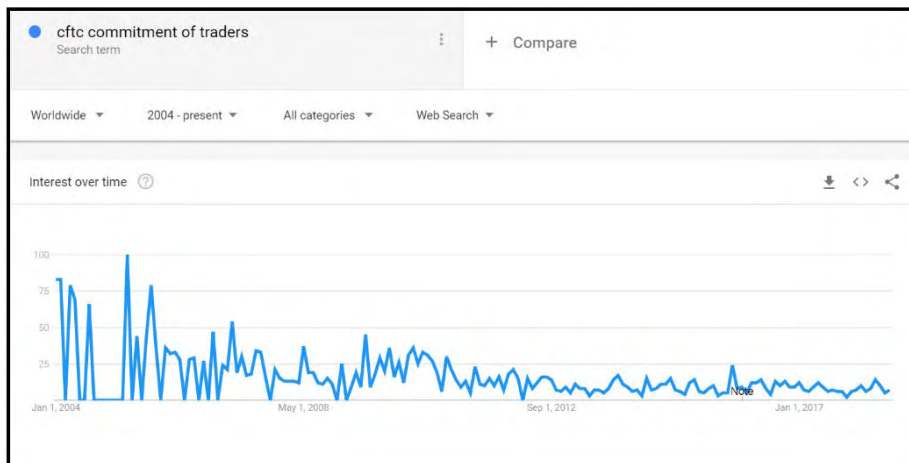


Illustration 6: Google Analytics CFTC Reports: Summary International Views.

there was less interest in Cotton on Call reports domestically, international interest was more frequent and Bank Participation report interest was not apparent.

Google Trends for
the COT reports

As shown on Illustration 7, available Google Trends data shows declining direct access to COT reports. While the numbers merely represents search interest relative to



the highest point on the chart, it provides context for the one CFTC Market reports most viewed by

Illustration 7: Commitment of Traders Searches From FY 2004 to November 2018

external audiences. However, we realize that third party vendors separately post data taken from the COT reports; we do not estimate third-party distribution here.

We also note that in the event of a government shutdown, cessation of the COT reports can generate interest and concern,¹⁸ which indicates there is still some value in the [COT reports](#) (or the information contained in them). Declining direct access on the www.cftc.gov website may indicate that individuals do not find direct access to the COT reports as useful as in the past (because they are static), or simply prefer third party¹⁹ availability/packaging of the same information (because it is interactive).

¹⁸ See, e.g., Bird, D. (2013, Oct. 4). Shutdown Shuts Down Commitment of Traders Reports From CFTC. *Wall Street Journal*, <https://blogs.wsj.com/moneybeat/2013/10/04/shutdown-shuts-down-commitment-of-traders-reports-from-cftc/>.

¹⁹ [CME Group – Commitment of Traders Tool](#) and [ICE Report Center](#).

Comments on CFTC.gov Large Trader and Cotton Reports

Between October 1, 2018 and July 30, 2018, CFTC received 248 inquiries related to either Large Trader (COT) or Cotton on Call reports. 78% of these comments were associated with report value.

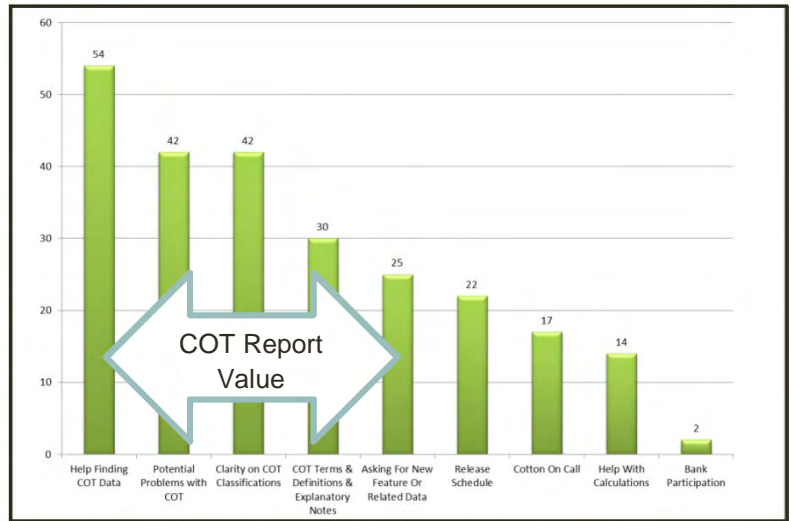
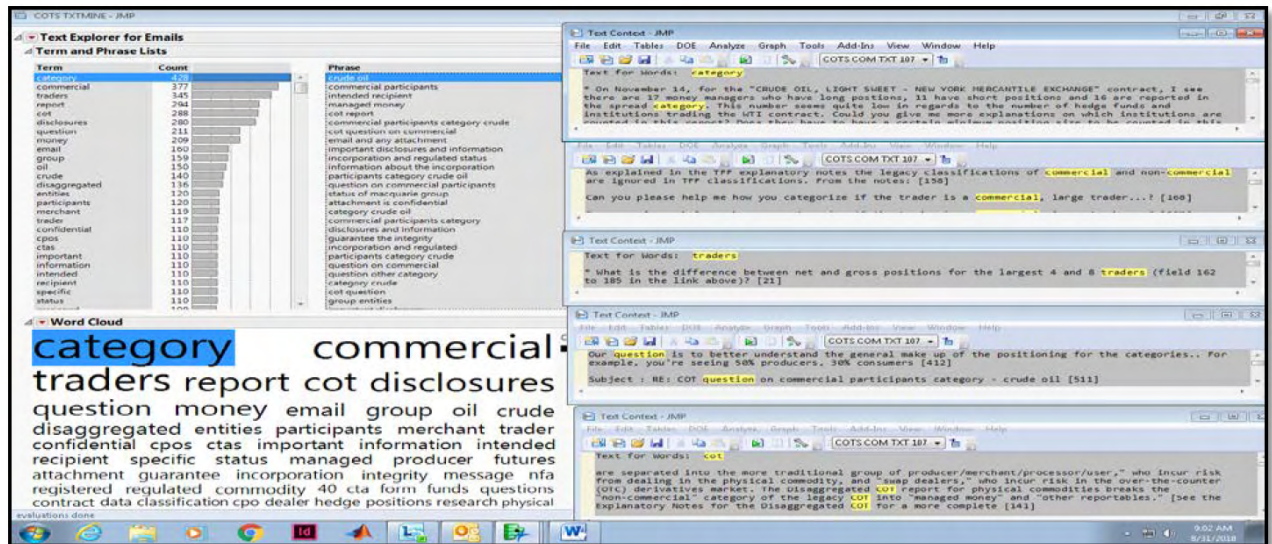


Illustration 8: Cotton on Call Report: Public Use Requests by Category.

COT Report Public Comments

Text analysis of 107 randomly selected emails to marketreports.cftc.gov showed most interest in the categorization or/classification and compilation of report data. To a lesser extent commenters also expressed concerns regarding the quality of reports.²⁰

Illustration 9: Word Analysis for Most Frequent COT Report Words Used by Stakeholders.

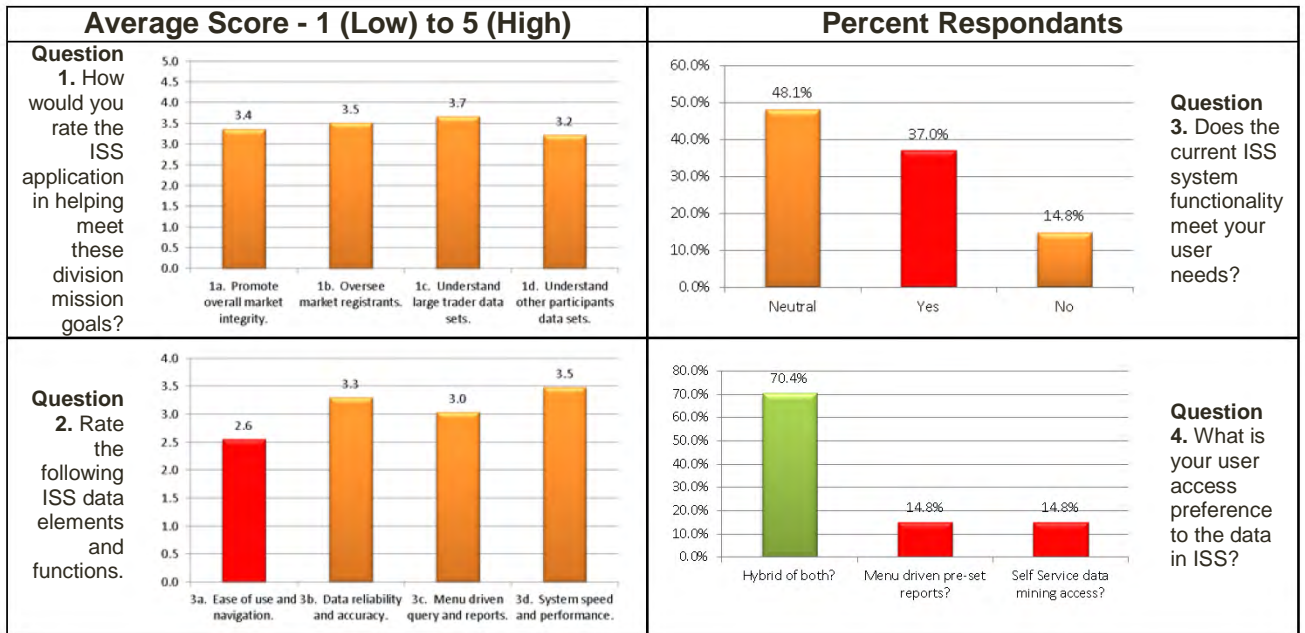


²⁰ We note that these emails were not solicited. In 2006, CFTC solicited comment on the COT reports. 71 FR 35627 (June 21, 2006). In response, CFTC received 4,659 comments from the U.S. and from 22 additional countries. At the time this set a record; the previous record was 1,062 comments received. CFTC, *Commission Actions in Response to the “Comprehensive Review of the Commitment of Traders Reporting Program”* (June 21, 2006) (<https://www.cftc.gov/sites/default/files/files/foia/comment06/foicf0603b002.pdf>).

Internal User Survey Results

We conducted a survey of ISS users. Respondents generally believed that ISS met division mission needs more so than not. However, functionality did not fully meet user expectations. While 23% of current users rated ease of use below average at 2.6 in a scale of one to five (1-5), 39% of current users rated data reliability and accuracy above average (3.3 out of 5), 39% of current users rated queries and reports just at average (3.0), and 42% of current users rated system performance above average (3.5). A clear majority of current users (70%) wanted both menu and self-service data access not currently available from the application.

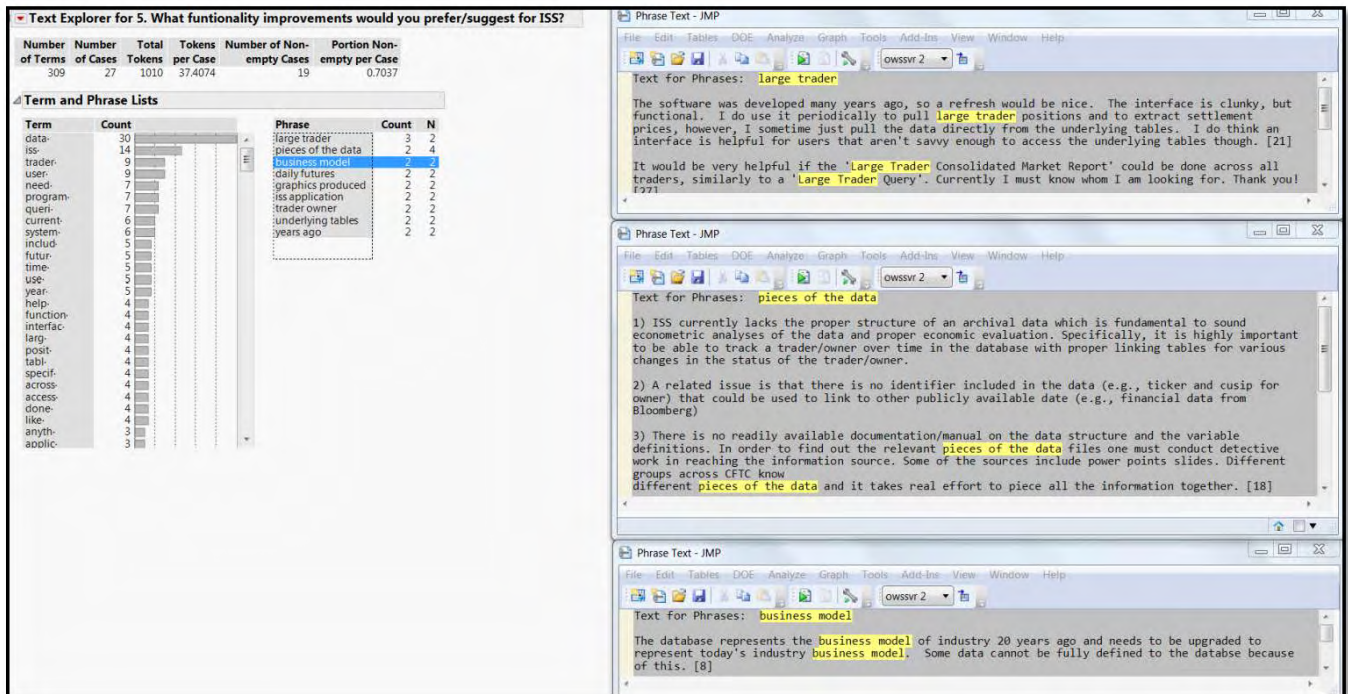
Illustration 10: Internal Survey Results – Questions 1, 2, 3, and 4.



Users also conveyed their desire for:

- A refresh or update to ISS’ “clunky interface;”
- The ability to generate reports across traders;
- Access to trend trader information; and
- Access to metadata information (that we noted earlier is specifically addressed under the Open Government Data Act).

Illustration 11: Internal Survey Results – Question 5: Suggested Improvements.



Report Substitutions Can Lessen Cost

Research of public reports using COT data identified commercial entities offering similar but more dynamic reports. For example, both CME Group (CME) and the Intercontinental Exchange (ICE) provide dynamic reporting of their respective COT data using queries and dashboards. In contrast to the static CFTC COT reports, ICE and CME users query data from a database. At CFTC, creating a customized static COT report can be costly because it must be done manually; a CFTC IT team cleans and prepares requested data for analysis in each instance. In contrast, the ICE and CME self-service tools already allow this functionality online. After an analyst connects to a raw data source, he or she can specify cleaning procedures and transformations that the data needs to go through before being presented in a CME or ICE dashboard.

Given the expense of creating static reports, the relatively low viewership of reports, declining COT report interest, and the desire for more report value, we note CFTC has an opportunity to reevaluate whether it is cost-effective to generate reports in its present form or to offer the technology advances of self-service tools that enhance customer experiences.

**U.S. Commodity Futures Trading Commission
Office of the Inspector General
Three Lafayette Centre
1155 21st Street, NW, Washington, DC 20581**

Recommendation

2. We recommend CFTC update business requirements for ISS and incorporate user and customer expectations in future ISS versions, if any, and as a part of an Enterprise-wide future-state roadmap that aligns with mission operations.

Appendix C

ETL Process for Data Staging

An ETL (Extract, Transform, and Load) process enables ODT to load data from source systems into the ISS database.²¹ An ETL tool is a means of taking data from one or many formats, transforming it and loading it into the database.

ETL Best Practices

Researching best ETL practices, we noted several common attributes across various platforms. Specifically:

- Modularity, that is, creating reusable code for ETL processes. ETL modularization helps avoid writing the same difficult code over and over, and reduces the total effort required to maintain the ETL architecture.
- Using ETL Staging Tables. Often, the use of interim staging tables can improve the performance and reduce the complexity of ETL processes.
- Error Handling - When suspect data is discovered, there needs to be a system for cleansing or otherwise managing nonconforming rows of data.
- Auditing & monitoring ETL jobs to ensure that the ETL jobs are performed as intended. Key attributes for these competencies includes, (1) logging, (2) checking for errors that also support auditing of row counts, financial amounts, and other metrics, and (3) data lineage, that is, documenting data source(s), when it was loaded, and how it was transformed.
- Using the appropriate ETL tool(s). There is a proper tool for every platform. However, email notifications in ETL processes add unnecessary complexity and potential failure points.

²¹ For background please see:
<http://www.dbta.com/Columns/SQL-Server-Drill-Down/Powerful-ETL-Technologies-in-the-Microsoft-Data-Platform-109419.aspx>
<https://aws.amazon.com/blogs/big-data/top-8-best-practices-for-high-performance-etl-processing-using-amazon-redshift/>
<https://www.timmitchell.net/etl-best-practices/>
<https://www.computerweekly.com/tip/Six-ETL-best-practices-followed-by-Shoppers-Stop>
<https://www.timmitchell.net/post/2017/06/14/etl-staging-tables/>.

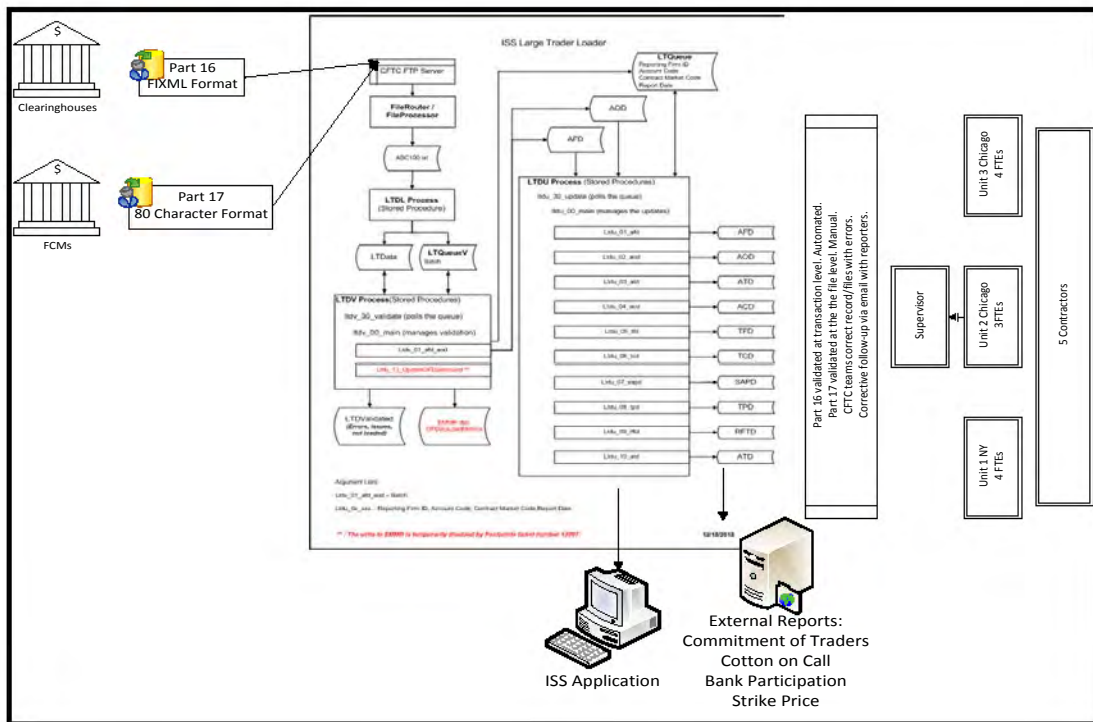
According to these practices, the current way of dealing with a failing ETL is to remedy any architectural and design mistakes by purchasing licenses and hardware. This approach may be merely disguising the symptoms, rather than tackling the root cause. Thus, CFTC may need to consider the current cost, effectiveness, and reliability of ISS data cleansing as used internally, and as the basis for CFTC's external market reports.

ISS ETL Process Can be Enhanced

When evaluating the ISS ETL process, we noted ODT employs certain best practices such as modularity, but could enhance ETL security, reliability, and efficiency by migrating to an updated platform.

As depicted in Illustration 12, ISS data is received from numerous traders via FTP; a 30 year old transfer protocol where security is not a chief benefit. In addition, FTP offers no traceability, that is, no way to see who accessed what information. This loophole has made it easy for cyber criminals to hack into FTP servers, retrieve shared information and leave without a trace.

Illustration 12: ISS Data ETL Process as Performed by CFTC.



As it relates to reliability, error handling is manually intensive because staff relies on email communication to correct submission errors. In our attempts to quantify errors in context of basic metrics such as row counts and financial amounts, we noted that logs for data load metrics have remained offline since 2017. The database manager explained that, due to the structure of the current data CFTC receives and the ISS system itself, the only way to determine rejected and unprocessed records and files is to forensically research each submission. There is a table that tracks errors on processed data, but it does not provide an accounting of whether or not the record that created the error condition was replaced. The system does not notify the reporter when a problem has occurred; it merely logs the fact that there was an error, and moves on. In other instances, if there is an error condition, the entire file is rejected, and the reporter is notified. The reporters know and understand that when this happens they are non-compliant with the rule until they re-submit. Thus, presently, ODT cannot easily determine whether the error conditions were corrected, or whether they still exist.

As it relates to efficiency, we noted from ISS logs that the current ETL configuration requires support from ODT personnel, contractors, and other CFTC staff in mission operations. More specifically, to service 55 front-end users, 33 support personnel²² (21 ODT) interfaced with ISS during the first half of calendar year 2018. This indicates that, for every 3 active ISS users, 2 support personnel are necessary to complete the ETL cycle and maintain the ISS. While we recognize this situation exists because the current configuration uses stored procedures to facilitate the transformation of files submitted, more modern platforms use staging areas and fully automate routines for bulk loading.

The current version of the ISS and databases are hosted on a Microsoft SQL Server 2008 platform. In April 2018, a change request was proposed to migrate the current platform to a 2014 cloud platform. We are encouraged by this step but note that Microsoft has since introduced powerful ETL features to the cloud via the Azure Data Factory.²³ Other vendors such as Amazon have similarly fielded solutions that facilitate and track ETL processes end-to-end securely. CFTC may wish to consider these options.

²² Defined as active users with Create, Read, Update, and Delete privileges.

²³ <https://docs.microsoft.com/en-us/azure/data-factory/introduction>.

Recommendation

3. Modernize the ISS to enhance the traceability, efficiency, and error handling of ETL processes, which will require a determination whether to update the ISS platform to achieve these goals or, based on a consideration of costs and benefits, to migrate to an updated platform.

Appendix D

Change Management Policy and Procedures

CFTC has implemented a Change Request Process²⁴ that requires approval from the change management board prior to making any system changes. This includes a security impact assessment.

The Change Control Board (CCB) is responsible for administering changes to configuration items (CI) in the enterprise Information Technology (IT) infrastructure of the CFTC. The CCB represents the interests of various CFTC divisions, offices and programs by enabling change decisions to be based on knowledge of change impact and benefits. It provides a clear and orderly process for tracking changes and for communicating information about change information to the various divisions and offices within the CFTC.

The CCB performs these tasks to:

- Ensure that proposed changes will not negatively impact current operations;
- Evaluate and approve, disapprove, or defer proposed changes;
- Review and authorize the establishment or changes to program baselines;
- Approve updated baselines and documentation; and
- Ensure the implementation of approved changes.

When analyzing the change management process for ISS 8.12.1,²⁵ we found the database team followed the ODT policy and procedures by submitting a standard change request with a security impact assessment for approval. Changes were tested for quality, a back out plan was documented, and the code changes were placed in production thereafter. Following the CCB process allowed ISS database managers to reduce the overall negative impact of system changes and to avoid compromising the CFTC IT enterprise infrastructure with harmful code.

Recommendation: None

²⁴ Change Control Board Guidelines and Process Revision 1.1.2, Apr. 06, 2017.

²⁵ The following changes are contained in this release: Users can cancel a trader and successfully move associated files. The requested changes are considered low risk.

Appendix E

Security and [REDACTED]

The CFTC operates a comprehensive system to collect information on market participants. Under CFTC's regulations, the Commission collects market data and position information from exchanges, clearing members, futures commission merchants (FCMs), foreign brokers, and traders. To ensure privacy of the information they provide, the CFTC has assigned confidential reporting numbers to reporting firms and traders. The Commission is prohibited under Section 8 of the Commodity Exchange Act, 7 USC 12, from publicly disclosing any person's positions, transactions, or trade secrets, except under limited circumstances. Under its large trader data program, the CFTC stores data collected in the ISS and has classified its contents as confidential information; including highly confidential trading information and sensitive personally identifiable information.

[REDACTED]

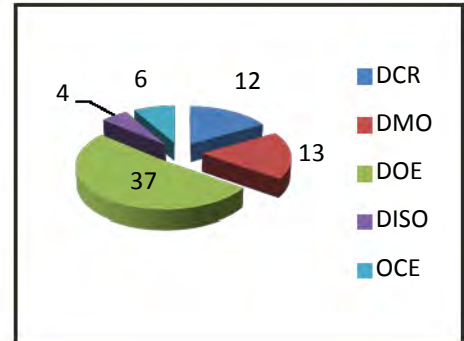
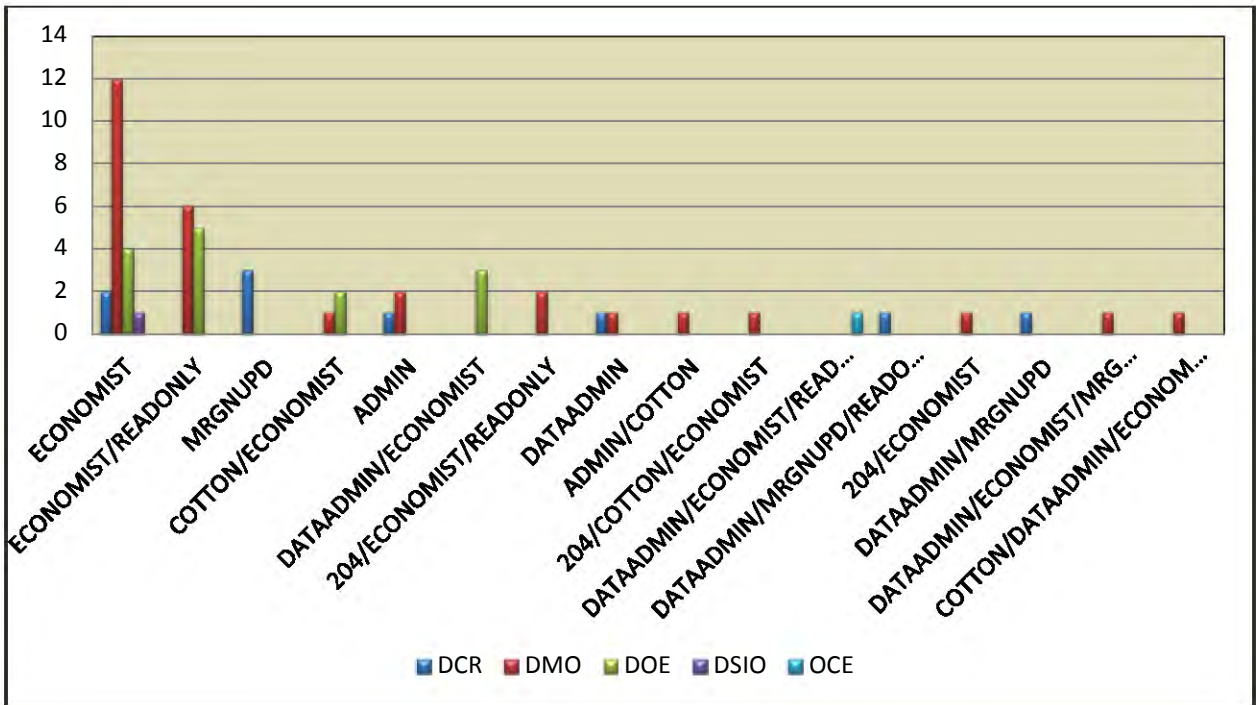


Illustration 13: ISS Users by CFTC Division.

[REDACTED]



Illustration 14: ISS Users Categorized by Roles.



System Limitation

ISS, as a legacy system, was designed prior to current security requirements and is not readily able to comply with the above requirements. Unless brought into compliance, the CFTC maintains the risk that confidential and sensitive market and PII data could be exfiltrated without authorization.

Recommendation

4. Given current federal security standards, we recommend CFTC re-evaluate [REDACTED] requirements for ISS.

We previously reported 2 other findings²⁷ related to the [REDACTED] – ISS [REDACTED]. We recommended that management improve [REDACTED]. Management agreed with our recommendation and is taking steps to remediate associated risks.

²⁷ CFTC'S Compliance with FISMA for Fiscal Year 2018, Report Number: 18-AU-08, October 30, 2018.



Appendix F

Security Concern for Similar Legacy Applications

“PowerBuilder” is a rapid application development tool for building and maintaining Windows applications, and provides core database functions. It has been in use since 1991, peaking around 1998. While PowerBuilder's market share has declined over the years, several CFTC applications, including the ISS, still use it today. Specifically, in addition to ISS, Master Data Services lists the following CFTC systems were built in PowerBuilder and still in use:

- ADD Content Manager - ADD Content Manager allows CFTC staff to review and approve content for CFTC.gov and manage the Content Refresh Notification Schedules. It has been identified as a low risk system.
- Commission and Staff Letters - Commission and Staff Letters (CSL) has been developed for the Division of Market Oversight to capture, manage, report, and publish No-Action, Exemption, and Interpretation actions initiated by both the Commission and by external registered and non-registered entities. The risk level is unknown.
- Filings and Actions – Filings and Actions (FILAC) allows CFTC staff to enter, modify, query, and publish submissions associated with organization, product, rules and foreign filings and actions. The risk level is unknown.

Name	Acronym	Risk Impact	Description	Platform	LastRelease
Commission and Staff Letters	CSL		Commission and Staff Letters (CSL) has been developed for the Division of Market Oversight to capture, manage, report, and publish No Action, Exemption, and	PowerBuilder	08/24/16
ADD Content Creator	ADD CWCC	Low	ADD Content Creator (ADD CWCC) compares pending content to current content on CFTC.gov. It identifies new content, modified content, and content that shou	PowerBuilder	05/09/18
ADD Content Refresh	ADD CWCR	Low	ADD Content Refresh (ADD CWCR) runs as a scheduled task each night. It replaces currently published content with pending content that has been approved an	PowerBuilder	05/09/18
ADD Content Manager	ADD CWCM	Low	ADD Content Manager allows CFTC staff to review and approve content for CFTC.gov and manage the Content Refresh Notification Schedules.	PowerBuilder	05/09/18
Correspondence Tracking System		Low	The Correspondence Tracking System tracks the Commission's controlled and uncontrolled correspondence. Receive incoming external correspondence such as e	PowerBuilder	05/01/15
Integrated Surveillance System	ISS		Integrated Surveillance System (ISS) collects futures and options end-of-day position data for large traders, and open interest, volume, price, and clearing mem	PowerBuilder	07/11/18
Filings and Actions	FILAC		Filings and Actions (FILAC) allows CFTC staff to enter, modify, query, and publish submissions associated with organization, product, rules and foreign filings and	PowerBuilder	09/21/16

Illustration 15: Legacy Applications Built on the PowerBuilder Platform.²⁸

While there are cybersecurity features for modern applications that allow compliance with federal security requirements, ██████████ for these other legacy applications may also pre-date current requirements for securing federal systems. Revaluating risk and security requirements for these legacy applications would reduce the opportunity for hackers to use old, trusted exploits. Many legacy applications also fail to document the changes made over time, leaving them vulnerable to systemic weaknesses, and gaps for malware. We recognize that legacy applications such as ISS remain in use because they are considered irreplaceable mission-critical systems with highly sensitive data; we believe this may result in a larger risk factor and tempting target for exfiltration.

Recommendation

5. We recommend CFTC review security risks of other legacy applications and assure compliance federal information security standards.

²⁸ Correspondence Tracking System – May have been retired with the deployment of CSL.

Appendix G

Background, Objective, Scope and Methodology

BACKGROUND

In recent years, Data Governance practices have become more defined within federal agencies, and requirements for effective information security programs have evolved. For example, the *President's Management Agenda* lays out a cross-agency goal to leverage data as a strategic asset. Achieving this goal requires data, accountability, and transparency initiatives to provide the tools to deliver visibly better results to the public, while improving accountability to taxpayers for sound fiscal stewardship and mission results. Investments in policy, people, and processes are key elements of this transformation and encompass all relevant governance, standards, and infrastructure and challenges of operating in a data-driven world. Executive Order 13800, *Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure* (May 11, 2017), also emphasizes safeguarding high-risk and value assets (including data). The Information Systems Audit and Control Association *Enabling Information* framework states that information or data, in context, should deliver value for its stakeholders, translating to achieving enterprise goals.²⁹

Data exists throughout enterprises; almost all stakeholders, processes and business activities rely on data at some level and to some degree. If data cannot be kept accurate, up to date, reliable, and secure, risk may increase across business, operational, and compliance domains. We note that a Data Governance program should ensure the following:

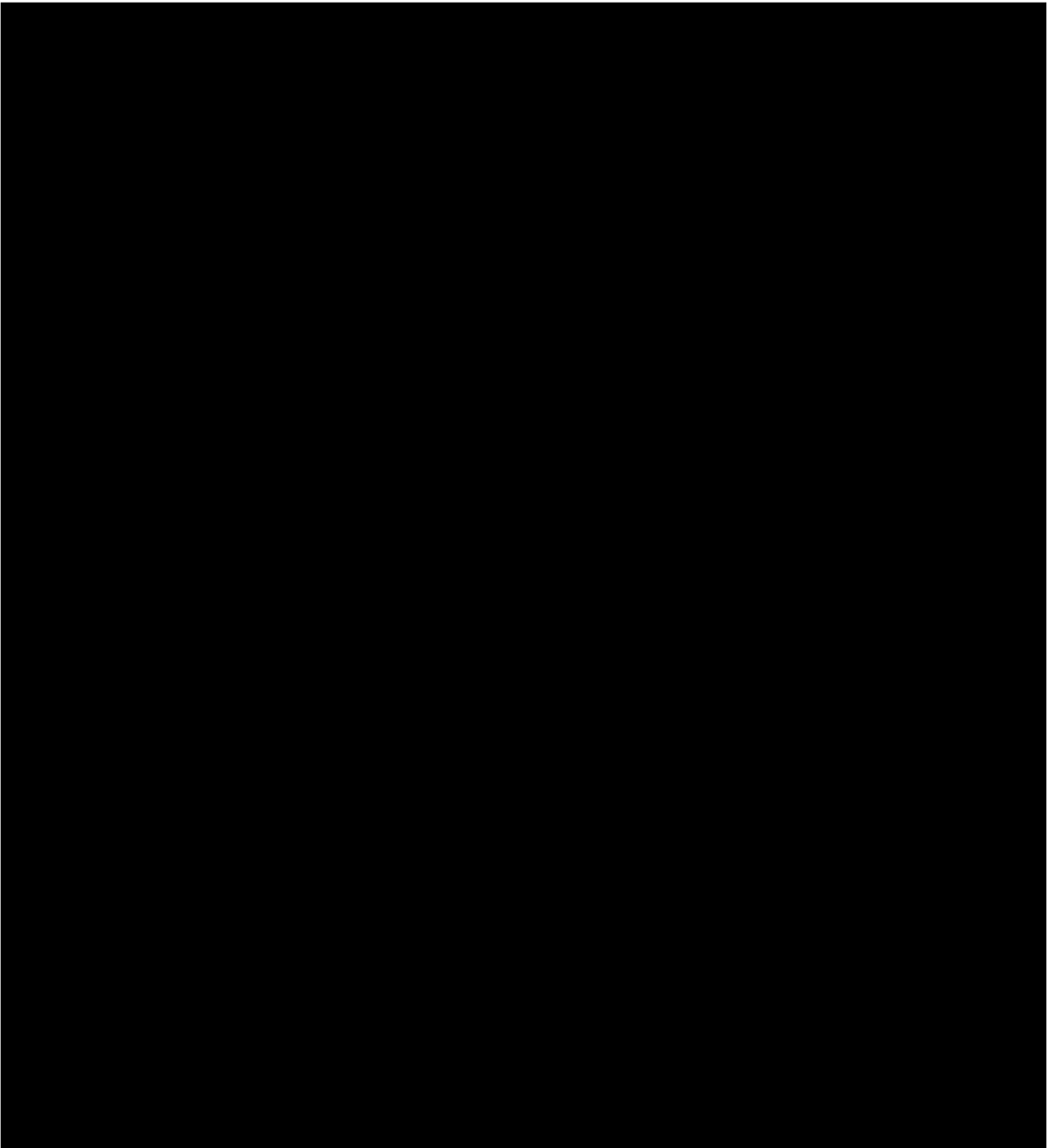
- Stakeholder needs, conditions and options are evaluated to determine balanced, mutually agreed enterprise objectives to be achieved through the acquisition and management of data/information resources;
- Direction is set for data/information management capabilities through prioritization and decision making; and
- Performance and compliance of data/information resources are monitored and evaluated relative to mutually agreed-upon (by all stakeholders) direction and objectives.

²⁹ https://m.isaca.org/COBIT/Documents/COBIT-5-Enabling-Information-Preview_res_Eng_0214.pdf.

In summary, a Data Governance program reflects the practice of evaluating requirements and bringing direction and control over data and information so that users have access to that data and can trust and rely on it.

At the CFTC, four groups are responsible for enabling information as a strategic asset; three are highlighted that have ISS responsibilities. As depict in Illustration 16, the Data Management Branch performs back-end database services, System and Services acts as a front-end bridge to stakeholders, and Policy and Planning ensures information security through a Chief Information Security Officer.

Illustration 16: Office of Data Technology (ODT) Organization by Branch.



OBJECTIVE, SCOPE, METHODOLOGY

The OIG reviewed CFTC's Data Governance program. Our objective was to assess the maturity CFTC's Data Governance practices using the Information Surveillance System (ISS); a system that hosts confidential and sensitive market data. Our scope evaluated practices for (1) defining business requirements, (2) extracting, transferring, and loading data, (3) managing changes, (4) maintaining stakeholder value, and (5) securing data.

To assess CFTC's Data Governance program maturity, we reviewed available documentation and methodologies, researched federal requirements and standards, and private organizations best practices. Relevant federal requirements reviewed included:

- The Clinger-Cohen Act, Pub. L. 104-106, Division E (Feb. 10, 1996).
- Executive Order 13800, Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure (May 11, 2017).
- Federal Information Technology Acquisition Reform Act, Pub.L. 113-291 (Dec. 19, 2014).
- OMB Circular A-11, Preparation, Submission and Execution of the Budget.
- OMB Circular A-130, Management of Federal Information Resources.
- E-Government Act of 2002, Pub. L. 107-347 (Dec. 17, 2002).
- Applicable NIST Guidance on Data Security and Access.

To evaluate practices, we analyzed business requirements for ISS focusing on required data formats, submission processes, and reported challenges, if any. To evaluate CFTC's extracting, transferring, and loading data (ETL) process, we evaluated schemas and procedures, applicable contracts, and interviewed responsible staff. Our review of change management practices covered policies and procedures, change requests, and approvals. We surveyed active CFTC business users to evaluate stakeholder use and perceived system value. We analyzed Google analytic statistics to assess external stakeholder value. Given ISS hosts high value assets, we evaluated system user activity logs, focusing on [REDACTED]. We examined ISS activity logs and inquired about their reliability. The system logs used were considered sufficiently reliable for the purpose of this review. We conducted our audit in accordance with *Generally Accepted Government Auditing Standards (GAGAS)*.

Appendix H

ISS Background and History

BACKGROUND

ISS organizes and stores end-of-day position reports electronically filed with the Commission by exchange clearing members, FCMs, Designated Contract Markets,³⁰ and foreign brokers (collectively called “reporting firms”). ISS allows the agency to monitor the daily activities of large traders, key price relationships, and relevant supply and demand factors.

ISS data enables the agency to analyze the composition of the market, such as the participation in the market by commercial versus non-commercial traders and the open interest held by certain occupational categories. ISS collects and processes daily futures and options position data for large traders from reporting firms and daily open interest, volume, price and clearing member data from exchanges. This data is supplemented by related cash market price data from a variety of sources. In addition, ISS receives and stores identifying information concerning each large trader and reportable account. ISS data is also used to provide public reports, such as the COT report and the Cotton on Call report.³¹

³⁰ CFTC states: Designated contract markets (DCMs) are exchanges that may list for trading futures or option contracts based on all types of commodities and that may allow access to their facilities by all types of traders, including retail customers.

<https://www.cftc.gov/IndustryOversight/TradingOrganizations/index.htm>.

³¹ CFTC Privacy Impact Assessment, Integrated Surveillance System (ISS), Sept. 30, 2014.

<https://www.cftc.gov/sites/default/files/idc/groups/public/@privacyoffice/documents/file/integratedsurveillancesystem.pdf>.

TIMELINE OF MAJOR EVENTS³²

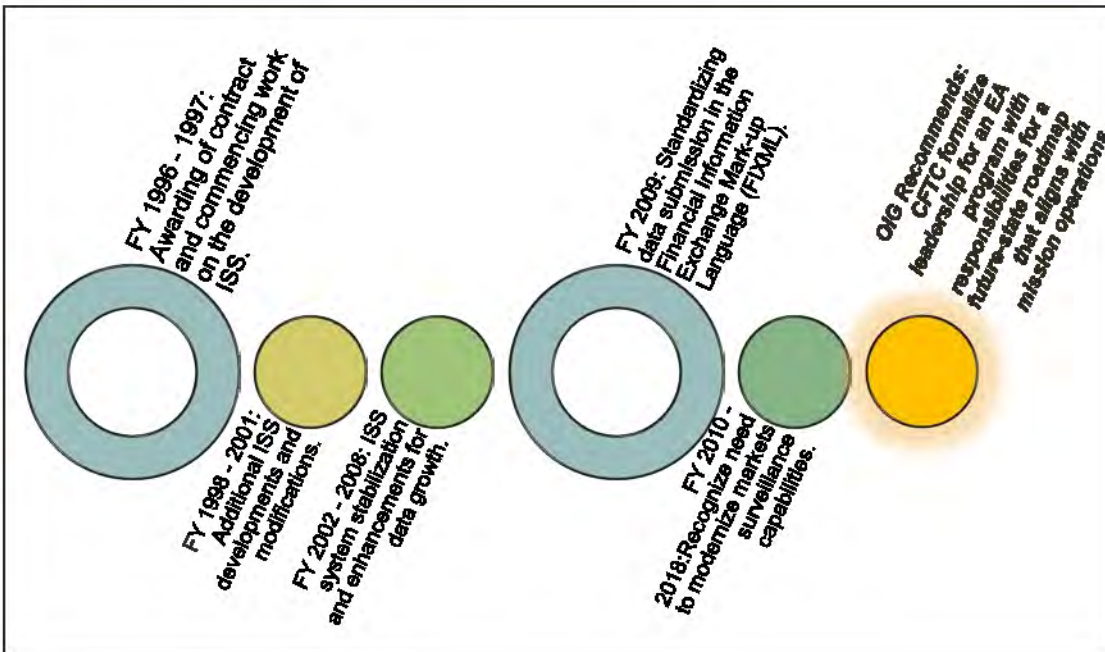


Illustration 17: ISS Timeline

Fiscal Year (FY) 1996: CFTC awarded a contract and commenced work on the development of a new Integrated Market Surveillance System. The system is designed using client-server architecture.

FY 1997: CFTC deployed ISS for the processing of futures and large trader data.

FY1998: Deployed capabilities to obtain large trader option position data on a daily basis.

FY 1999: Modified the market surveillance system to use information from FCMs on the option positions of large traders.

FY 2000: Reengineered the market surveillance system based upon client-server architecture.

FY 2001: Contracted to provide software development support and enhancements to stabilize the ISS.

FY 2003: Transferred all Large Trader Reports into the ISS.

³² **Source:** Historic CFTC Annual Reports, Performance and Accountability Reports, and Presidents Budget. All reports are available here: <https://www.cftc.gov/About/CFTCReports/index.htm>.

FY 2004: Enhanced ISS to address changes and growth in the futures industry. Those changes included accepting markedly different contract markets.

FY 2005: Enhanced ISS to address changes and growth in the futures industry. Those changes included the automation of the collection and review of data from exempt commercial markets.

FY 2008: Enhanced ISS to better display positions in futures months.

FY 2009: Standardized data submission in the Financial Information Exchange Mark-up Language (FIXML).

FY 2010: Invested in automated alerting and workflow and position limit monitoring changes to adapt to market changes.

FY 2011-2015: Communicated plans to modernize this mission critical system, which was originally developed when manual detection methods were sufficient to monitor market conditions. The intent was to enable focus on greater market manipulation detection and reporting, increased internal and external data transparency enhanced reports, improved case management to track market surveillance activities, integration of large trader data with intraday trading activity, improving the automated collection of data from industry participants, and enhancing capabilities to manage position limits.


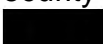
FY 2016 – 2017: CFTC did not specifically discuss ISS in its President’s Budget & Performance Plan. However, CFTC requested funding to enhance surveillance capabilities to keep pace with the increasing technological sophistication of the markets—in particular, the increasing use of automated trading.

FY 2018: CFTC stated that information technology services support data acquisition and analysis that are crucial to conducting effective surveillance and oversight of an increasingly diverse electronic marketplace, detecting/investigating illegal activities, and conducting regulatory oversight of Commission registrants. In support of the surveillance function, the Commission planned to improve the quality of data ingested using a multifaceted approach, including development of detailed data quality, acceptance specifications/standards, data validation checks, and data quality monitoring and reporting.

CFTC recognized that in the modern marketplace, where automated trading dominates, many manipulation strategies cannot be determined by using legacy data already filed with the Commission by market participants or available in the market.

Appendix I

Notice of Findings and Recommendations

FY 2019 Finding(s):	FY 2019 Recommendation(s):	Management Concurrency (Y/N)
NFR-DGA-ISS-01-2019 Deficiency in maturing an enterprise Data Governance Program.	We recommend CFTC: 1. Set a timeframe to fully implement plans for its Data Governance framework and, if not already, synchronize with goals outlined in the Federal Data Strategy and Open Data Government Act requirements.	Yes
NFR-NGA-ISS-02-2019 Deficiency in data management controls for ISS.	2. Update business requirements for ISS and incorporate stakeholder expectations in future ISS versions as a part of an Enterprise Architecture program plan that aligns with mission operations.	Yes
	3. Modernize the ISS to enhance the traceability, efficiency, and error handling of ETL processes, which will require a determination whether to update the ISS platform to achieve these goals or, based on a consideration of costs and benefits, to migrate to an updated platform.	Yes
	4. Given current federal security standards, re-evaluate  requirements for ISS.	Yes
	5. Review security risks of other legacy systems and assure compliance with information security standards.	Yes

Appendix J

Management's Comments



U.S. COMMODITY FUTURES TRADING COMMISSION
Three Lafayette Centre
1155 21st Street, NW, Washington, DC 20581
Telephone: (202) 418-5000
Facsimile: (202) 418-5521
www.cftc.gov

Office of Data & Technology

MEMORANDUM

TO: Miguel A. Castillo, CPA, CRMA,
Assistant Inspector General for Audits

FROM: John L. Rogers, Chief Information Officer

DATE: May 2, 2019

SUBJECT: Commodity Futures Trading Commission Management Responses to
the Review of CFTC's Data Governance Program: Integrated
Surveillance System

We appreciate the opportunity to respond to the subject draft report. The Office of Data and Technology (ODT) within the CFTC concurs with the audit's recommendations. ODT is committed to working with the appropriate stakeholders to address the issues and concerns mentioned in the referenced report. ODT's management responses and actions taken are listed below for each recommendation:

- 1. Set a timeframe to fully implement plans for its Data Governance framework and, if not already, synchronize with the goals outlined in the Federal Data Strategy and Open Data Government Act requirements;*

The Office of Data and Technology (ODT) concurs with the Office of the Inspector General's assessment and their recommendation.

ODT is working on formulating a Data Governance framework that aligns with the goals outlined in the Federal Data Strategy and Open Data Government Act requirements. Subject to funding, a full implementation plan will be developed and executed to implement the first level of policies, procedures, and governing bodies based on the framework.

- 2. Update business requirements for ISS and incorporate stakeholder expectations in future ISS versions as a part of an Enterprise Architecture that aligns with mission operations;*

ODT concurs with the Office of the Inspector General's assessment and their recommendation.

ODT is working on a plan to address the concerns and recommendations in the OIG report. The plan will include new requirements gathering efforts, meeting with users of the Integrated

Surveillance System (ISS) from the Division of Market Oversight, the Division of Enforcement, the Division of Clearing and Risk, the Division of Swaps and Intermediary Oversight, the Office of the Chief Economist, and the Market Data Operations section of ODT. ODT will ensure changes will align with mission operations and federal information security standards.

- 3. Modernize the ISS to enhance the traceability, efficiency, and error handling of ETL processes, which will require a determination whether to update the ISS platform to achieve these goals or, based on a consideration of costs and benefits, to migrate to an updated platform;*

ODT concurs with the finding that it is appropriate to modernize the ETL processes that support ISS data loading.

ODT has developed and is promulgating an upgraded standard for data transmission of ISS data. Along with the updated data standard, ODT will be upgrading the ETL process for ISS data. The new ETL process will meet all of the ETL best practices as recommended in the report. A main principle of the improved ETL process would be to automate the data validation and verification process such that data reporters will be notified of issues with the data without the intervention of a CFTC staff member.

- 4. Given current federal security standards, re-evaluate [REDACTED] requirements for ISS;*

ODT concurs with the Office of the Inspector General's assessment and their recommendation.

As part of the requirements update and validation of ISS, ODT will ensure compliance with the current federal security standards.

- 5. Review security risks of other legacy applications and assure compliance with federal information security standards.*

ODT concurs with the Office of the Inspector General's assessment and their recommendation.

ODT will continue to enhance their compliance with the FISMA based on the NIST Risk Management Framework (RMF) process and Cybersecurity Framework as directed by OMB and Congress. Continuous monitoring efforts are underway and subject to resource constraints. ODT will continue to work to ensure full compliance with Federal information security standards.

If you require further assistance, please contact Naeem Musa, Deputy Director of Policy and Planning, at (202) 418-5485.



1155 21st Street N.W.
Washington, DC 20581
202 418-5100