DATA STRATEGY

FISCAL YEARS 2022–2026

▸ Enhance use of data assets

▸ Design and build integrated data infrastructure

▸ Promote data awareness and literacy

▸ Strengthen data management and governance

GOAL
TIME CREATIVITY
SIGN STRATEGY TARGET
GROWTH INSPIRATION
MARKET IDEAS
TEAM
SOLUTION INNOVATION
TEAMWORK VISION
ANALYSIS
PROGRESS
RISK

Contents

About the SEC Data Strategy

The SEC's Office of the Chief Data Officer (OCDO) developed this Data Strategy with advice from the SEC's Data Management Board. The goals and initiatives guide the agency's Data Management Program[1] and support the Commission's Strategic Plan for Fiscal Years 2022 through 2026.[2]

The Data Management Program includes enterprise-level data management projects and responsibilities and has the following overall objectives: enabling collaborative management of data as a strategic enterprise asset; ensuring secure, efficient, and effective collection, management, use, protection, dissemination, and generation of data within the SEC; and ensuring, to the extent practicable, that the SEC maximizes the value and use of data to carry out its mission.

Data Vision

The SEC enables high-quality staff data analysis, provides well-structured data to the public, and manages data as a strategic asset.

Figure 1: Goals and Initiatives Summary

Goals and Initiatives SUMMARY

Goal 1
Enhance use of data assets

Goal 2
Design and build integrated data infrastructure

Goal 3
Promote data awareness and literacy

Goal 4
Strengthen data management and governance

"

The availability and accuracy of data that the agency generates, collects, and releases has always been viewed as a defining feature of U.S. capital markets that inspires investor trust and confidence.

Austin Gerig
Chief Data Officer

"

Message from the Chief Data Officer

AUSTIN GERIG
CHIEF DATA OFFICER

Data has always been central to the SEC's regulatory mission. For example, public companies must disclose to investors material and accurate information regarding their operations, financial condition, and management; and brokers, dealers, and other intermediaries must keep and preserve business records and records concerning securities transactions. The availability and accuracy of data that the agency generates, collects, and releases has always been viewed as a defining feature of U.S. capital markets that inspires investor trust and confidence. In this sense, the agency's data constitutes a public good that we must carefully protect and manage.

Title II of the Foundations for Evidence-Based Policymaking Act of 2018,[3] also known as the Open, Public, Electronic, and Necessary (OPEN) Government Data Act, sets forth a framework for managing government information as a strategic asset at each stage of the information's life cycle to promote accessibility and usability. This framework has accelerated ongoing data-related initiatives and provided another impetus to meet the challenges of today's increasingly complex, fast-paced, and interconnected markets. The OPEN Government Data Act makes agency chief data officers responsible for managing data assets;[4] improving the agency's infrastructure for data asset accessibility; and coordinating with others in the agency responsible for using, protecting, disseminating, and generating data to ensure that the data needs of the agency are met.

This Data Strategy establishes an agenda for implementing certain OPEN Government Data Act requirements and improving the management and use of data at the SEC. The goals described are foundational and interrelated, but are also designed to allow priorities to shift in response to changing market trends and the agency's evolving needs. Our embrace of agile project management and our use case-driven approach to investments will help ensure our efforts remain focused on value-generating activities.

It is my honor to serve as the SEC's first Chief Data Officer and present this Data Strategy. I look forward to advancing the Data Strategy's goals in close partnership with all of the agency's divisions and offices.
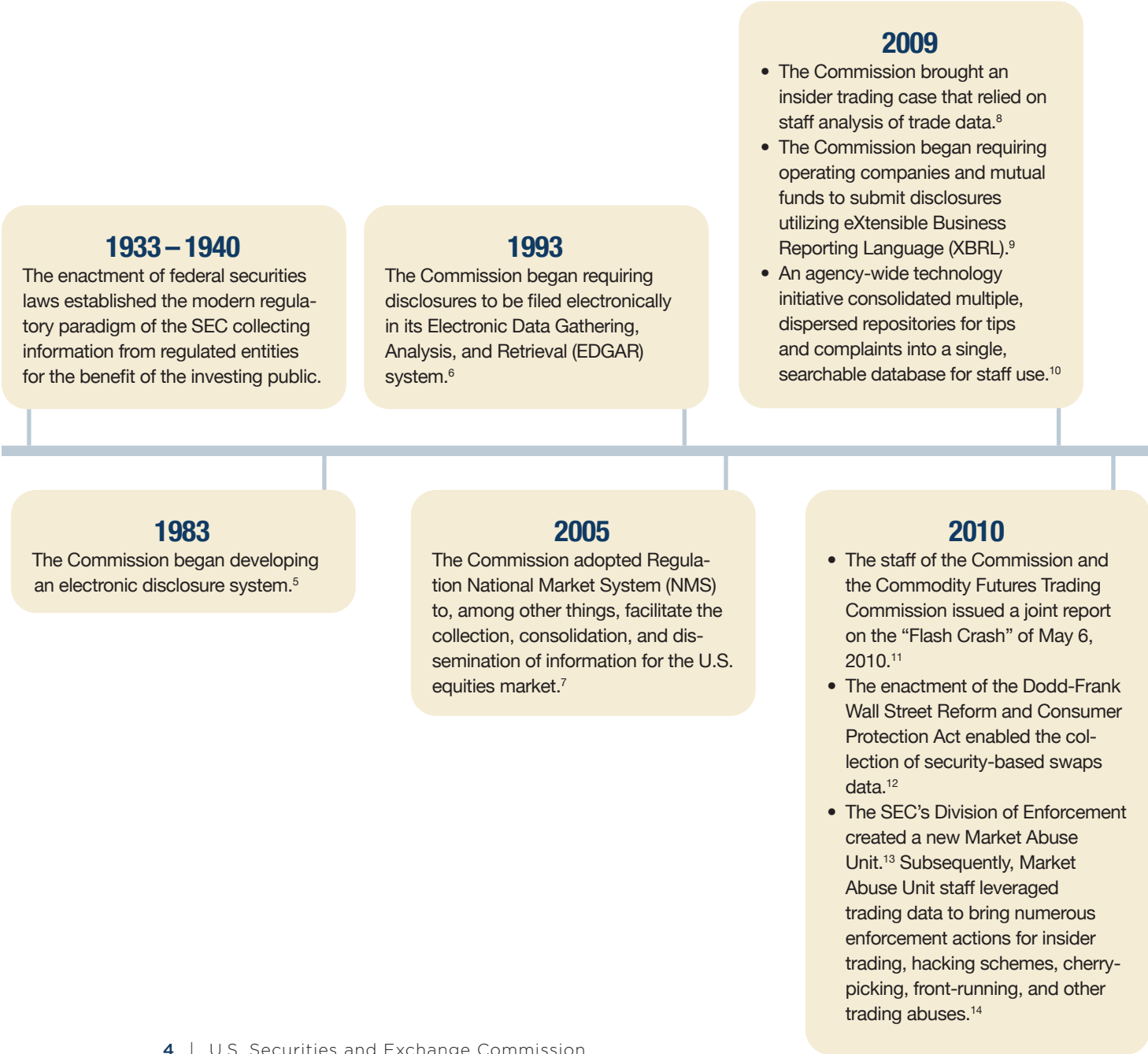
Austin Gerig
Chief Data Officer

The Vital Role of Data to the SEC Mission

Important developments in the creation of core data assets, in the implementation of new technologies that enabled more effective transmission or analysis of data, and in uses of data to enhance protection of investors or improve policy stand out as key milestones that have shaped SEC efforts to accomplish its regulatory mission.

Figure 2: Data-Related Milestones

1933 – 1940

The enactment of federal securities laws established the modern regulatory paradigm of the SEC collecting information from regulated entities for the benefit of the investing public.

1993

The Commission began requiring disclosures to be filed electronically in its Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system.[6]

2009

- The Commission brought an insider trading case that relied on staff analysis of trade data.[8]
- The Commission began requiring operating companies and mutual funds to submit disclosures utilizing eXtensible Business Reporting Language (XBRL).[9]
- An agency-wide technology initiative consolidated multiple, dispersed repositories for tips and complaints into a single, searchable database for staff use.[10]

1983

The Commission began developing an electronic disclosure system.[5]

2005

The Commission adopted Regulation National Market System (NMS) to, among other things, facilitate the collection, consolidation, and dissemination of information for the U.S. equities market.[7]

2010

- The staff of the Commission and the Commodity Futures Trading Commission issued a joint report on the "Flash Crash" of May 6, 2010.[11]
- The enactment of the Dodd-Frank Wall Street Reform and Consumer Protection Act enabled the collection of security-based swaps data.[12]
- The SEC's Division of Enforcement created a new Market Abuse Unit.[13] Subsequently, Market Abuse Unit staff leveraged trading data to bring numerous enforcement actions for insider trading, hacking schemes, cherry-picking, front-running, and other trading abuses.[14]

2022

The Commission brought a front-running case that was uncovered by staff analysis of Consolidated Audit Trail data.[24]

2016

- New data analytics, including the use of artificial intelligence, machine learning, and natural language processing techniques, were used to identify registrants with elevated risk profiles,[17] and a new exam analytics tool strengthened staff analysis of registrant trading activities.[18]
- The Commission adopted reforms to modernize and enhance the reporting and disclosure of information by registered investment companies and to enhance liquidity risk management by open-end funds, including mutual funds and exchange-traded funds.[19]

2020

- The SEC's Office of the Chief Data Officer was created.[21]
- New data on private funds (e.g., hedge funds and private equity funds) collected through Form PF enabled risk-targeted SEC examinations of private fund advisers and addressed a data gap in the Financial Stability Oversight Council's assessment of systemic risk in the U.S. financial markets.[22]

2012

The Commission adopted Regulation NMS Rule 613, requiring self-regulatory organizations to create the Consolidated Audit Trail.[15] The Consolidated Audit Trail tracks orders throughout their life cycle, providing a comprehensive view of orders and executions across U.S. markets.

2013

The agency implemented a Market Information Data Analytics System (MIDAS), which allows for the efficient collection and analysis of equity and futures order book data.[16]

2018

The Commission's adoption of amendments requiring the use of the Inline format for operating company financial statement information and mutual fund risk/return summary information enhanced the utility of that information.[20]

2021

The Consolidated Audit Trail data facilitated the agency's inquiry into the extraordinary market volatility in early 2021.[23]

Data-Related Goals and Initiatives

GOAL 1. Enhance use of data assets

Emphasize accessibility, structuring, and use of data assets to strengthen staff analysis and drive value for the public.

1.1 Maximize SEC use of core data assets[25]

Enhance data tools, infrastructure, and controls to enable faster and more cost-effective data analysis to assess market developments and risks, prioritizing uses of data assets that have the greatest potential to advance the SEC's regulatory mission.

1.2 Enable new combinations of data to facilitate discovery of vital regulatory insights

Enhance data tools, infrastructure, standards, and controls to enable staff to combine data for analysis to more effectively gain regulatory insights and prevent, detect, and enforce against improper behavior.

1.3 Increase accessibility and usability of disclosure data in electronic filings

Enhance the accessibility and usability of public disclosures in electronic filings by improving mechanisms for ascertaining needs of data users (both outside and inside the SEC) and translating feedback regarding data quality into work product that informs recommendations for future rulemaking.

GOAL 2. Design and build integrated data infrastructure

Create a data environment that supports a seamless user experience that is governed by consistent standards, enables collaborative analysis and insight sharing, integrates services from multiple providers as needed, and enhances data security.

2.1 Promote seamless integration of data and analytics services

Recommend data architecture for an enterprise data infrastructure to support analytics platforms and to facilitate collaborative data analysis between and within divisions and offices. Leverage cloud computing capabilities to manage large-scale data assets more efficiently and to facilitate collaboration.

2.2. Establish and apply uniform data and data architecture[26] standards

Establish and apply data and data architecture standards consistently across all data assets to facilitate work across data assets, reduce data siloes, enhance collaboration, facilitate blending of different data assets, and streamline data management.

2.3 Implement enhanced data security controls

Automate data categorization and security responses, focusing on tagging and managing access to sensitive information, to support data governance, compliance with government mandates,[27] and effective monitoring of data usage.

Goal 3. Promote data awareness and literacy

Increase use of analytical tools with a focus on the needs of SEC staff. Build a data culture characterized by widespread data literacy and data-driven decision-making, including through workforce training on data approaches and techniques.

3.1 Promote a data-driven culture

Carry out an enterprise-wide change management and communications plan to strengthen staff commitment to data hygiene best practices. A data culture in which every staff member recognizes the importance of being a responsible steward of data and practicing data-driven decision-making will help the SEC fully leverage data as a strategic asset.

3.2 Expand the SEC's internal data catalog to promote data awareness and facilitate collaboration

Augment the SEC's data catalog as a means of promoting change management through community building around increased data awareness and literacy. Use the catalog to strengthen the staff's understanding of data management best practices, promote data-driven decision-making, and help scale capabilities and talent across the Commission.

Goal 4. Strengthen data management and governance

Design and implement use case-oriented data management practices that balance security and usability, facilitate collaboration, enhance analysis, and improve the quality of SEC data.

4.1 Enhance data policies

Enhance existing data policies and implement new policies to simplify data governance across the Commission and strengthen data protection. Address categorization of data, metadata capture, and governance of data access and sharing.

4.2 Institute an enterprise-wide data quality program

Institute SEC-wide data quality processes to identify, catalog, and remediate data quality issues across data assets. Assign staff members responsibility for defining quality standards (based on industry standards, where possible) and developing plans for remediating identified quality issues. Include data quality descriptions in catalog records.
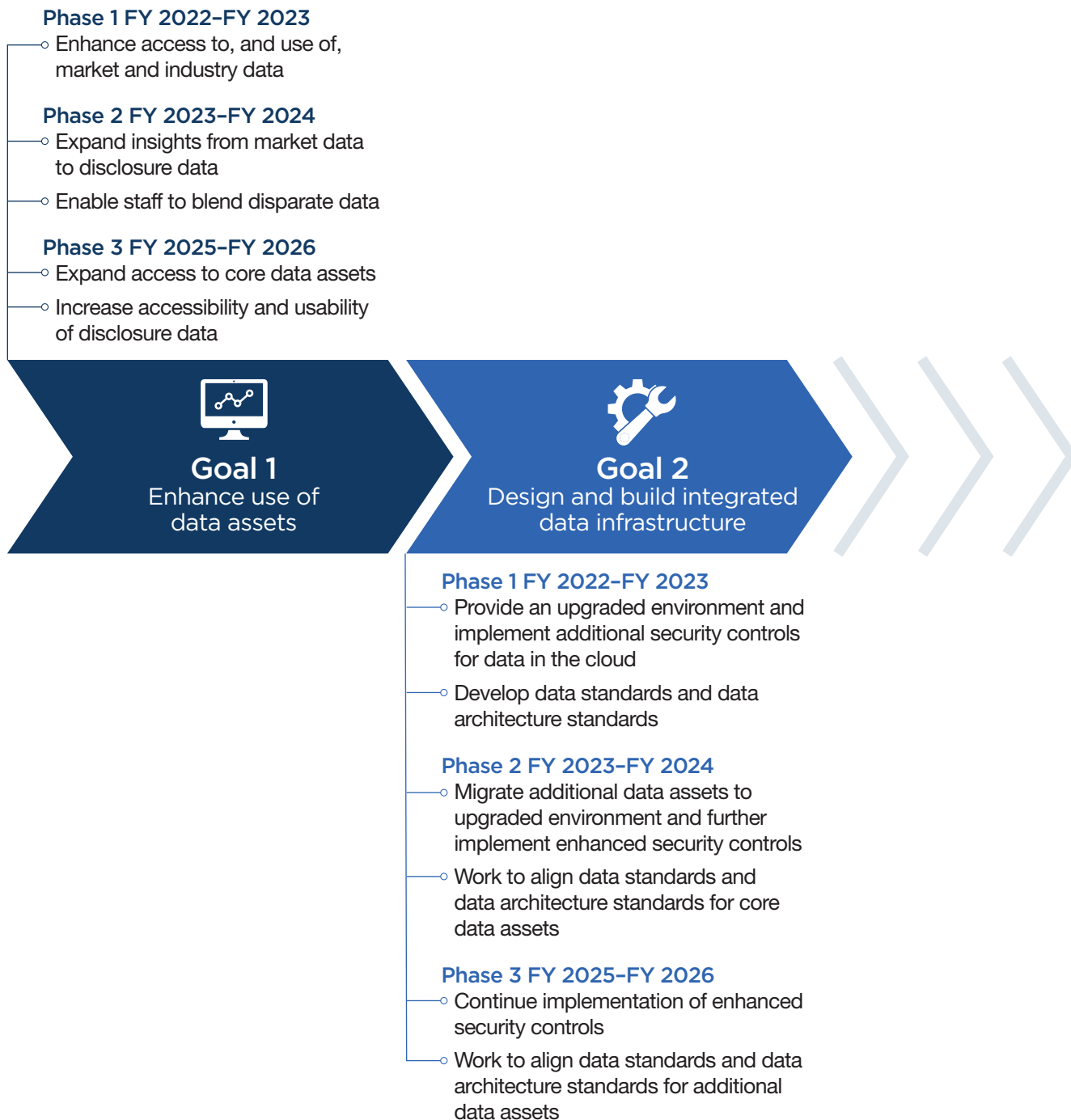
4.3 Establish an evaluation framework for data assets, infrastructure, uses, and policies

Develop a framework to assess the value, cost, and risk of agency data acquisition and retention plans, data services infrastructure, data usage, and data controls, policies, and procedures.

Strategy Roadmap

Following publication of this Data Strategy, OCDO will work with agency stakeholders to develop an implementation plan in accordance with agency resourcing decisions.

Figure 3: Strategy Roadmap

Phase 1 FY 2022–FY 2023
○ Enhance access to, and use of, market and industry data

Phase 2 FY 2023–FY 2024
○ Expand insights from market data to disclosure data
○ Enable staff to blend disparate data

Phase 3 FY 2025–FY 2026
○ Expand access to core data assets
○ Increase accessibility and usability of disclosure data

Goal 1
Enhance use of data assets

Goal 2
Design and build integrated data infrastructure

Phase 1 FY 2022–FY 2023
○ Provide an upgraded environment and implement additional security controls for data in the cloud
○ Develop data standards and data architecture standards

Phase 2 FY 2023–FY 2024
○ Migrate additional data assets to upgraded environment and further implement enhanced security controls
○ Work to align data standards and data architecture standards for core data assets

Phase 3 FY 2025–FY 2026
○ Continue implementation of enhanced security controls
○ Work to align data standards and data architecture standards for additional data assets

Phase 1 FY 2022–FY 2023
- Expand the data user community
- Promote collaboration among divisions and offices

Phase 2 FY 2023–FY 2024
- Enhance staff data literacy and analytical capabilities through training

Phase 3 FY 2025–FY 2026
- Integrate data literacy and analytical skills throughout all divisions and offices

Goal 3
Promote data awareness and literacy

Goal 4
Strengthen data management and governance

Phase 1 FY 2022–FY 2023
- Establish data policies
- Apply data policies to core data assets

Phase 2 FY 2023–FY 2024
- Apply data policies to core data assets
- Institute proactive data quality management for core data assets

Phase 3 FY 2025–FY 2026
- Apply data policies to additional data assets
- Institute proactive data quality management for additional data assets

ENDNOTES

1 The Data Management Program encompasses enterprise-level data management projects and responsibilities. The Data Management Board provides advice and recommendations on oversight of the Data Management Program. The Board is composed of senior officials from across the SEC who have authority over data management and use, and who have insight into alignment between the SEC's mission and its data management needs. *See* Sec. & Exch. Comm'n, SEC Data Management Board Charter (Nov. 2, 2021) (last visited Dec. 3, 2022). The Data Management Program does not encompass data-related initiatives of individual divisions and offices.

2 The Commission's Strategic Plan for Fiscal Years 2022 through 2026 has three primary goals: (1) protect the investing public against fraud, manipulation, and misconduct; (2) develop and implement a robust regulatory framework that keeps pace with evolving markets, business models, and technologies; and (3) support a skilled workforce that is diverse, equitable, and inclusive and is fully equipped to advance agency objectives. The Commission developed these goals to advance the SEC's mission, which is to protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation. *See* Press Release, Sec. & Exch. Comm'n, SEC Publishes FY22-26 Strategic Plan (Nov. 23, 2022).

3 *See* Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5529 (Jan. 14, 2019).

4 The OPEN Government Data Act defines the term "data" to mean "recorded information, regardless of form or the media on which the data is recorded" and the term "data asset" to mean "a collection of data elements or data sets that may be grouped together." 44 U.S.C. 3502(16) and (17).

5 *See* Sec. & Exch. Comm'n, *Electronic Filing and the EDGAR System: A Regulatory Overview* (last visited Dec. 3, 2022).

6 *See* id.

7 *See* Press Release, Sec. & Exch. Comm'n, SEC Adopts Regulation NMS and Provisions Regarding Investment Advisers Act of 1940 (Apr. 7, 2005).

8 *See* Press Release, Sec. & Exch. Comm'n, SEC Charges Wall Street Professionals and Others in Insider Trading Ring (Feb. 5, 2009).

9 *See* Sec. & Exch. Comm'n, *Structured Disclosure at the SEC: History and Rulemaking* (last visited Dec. 3, 2022).

10 *See* Sec. & Exch. Comm'n *Fiscal Year 2010 Performance and Accountability Report* (Nov. 15, 2010) at 11.

11 *See* Sec. & Exch. Comm'n, *Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues* (Sep. 30, 2010).

12 *See* Title VII of the Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376 (Jul. 21, 2010).

13 *See* Sec. & Exch. Comm'n, *Fiscal Year 2010 Performance and Accountability Report* (Nov. 15, 2010) at 10.

14 *See*, *e.g.*, Press Release, Sec. & Exch. Comm'n, SEC Charges Investment Advisers with Cherry Picking, Obtains Asset Freeze (Jun. 23, 2021); Press Release, Sec. & Exch. Comm'n, SEC Charges Investment Adviser and Its COO with Defrauding Clients (Jun. 29, 2021); Press Release, Sec. & Exch. Comm'n, SEC Charges Hedge Fund Trader in Lucrative Front-Running Scheme (Jul. 2, 2021); Press Release, Sec. & Exch. Comm'n, SEC Charges Netflix Insider Trading Ring (Aug. 25, 2021); Press Release, Sec. & Exch. Comm'n, SEC Charges Quant Analyst in Multimillion Dollar Front-Running Scheme (Sep. 23, 2021).

15 *See* Press Release, Sec. & Exch. Comm'n, SEC Approves New Rule Requiring Consolidated Audit Trail to Monitor and Analyze Trading Activity (Jul. 11, 2012).

16 *See* Sec. & Exch. Comm'n, *MIDAS: Market Information Data Analytics System* (last visited Dec. 3, 2022); Sec. & Exch. Comm'n, *Agency Financial Report Fiscal Year 2014* at 14, 18 (reporting the development of MIDAS to collect and analyze market data, permitting staff to better use data to monitor the integrity of markets); Gregg E. Berman, Associate Director, Off. of Analytics and Rsch., Div. of Trading and Markets, Sec. & Exch. Comm'n, *Address at SIFMA TECH Conference: Transformational Technologies, Market Structure, and the SEC* (Jun. 18, 2013) (describing the original impetus of MIDAS as to be able to more readily reconstruct market-wide events like the Flash Crash).

17 *See* Scott W. Bauguess, Deputy Director and Deputy Chief Economist, Div. of Econ. and Risk Analysis, Sec. & Exch. Comm'n, *Address at American Accounting Association Midwest Region Meeting: Has Big Data Made Us Lazy?* (Oct. 21, 2016).

18 *See* Sec. & Exch. Comm'n, Off. Of Compliance Inspections And Examinations, *National Exam Program Examination Priorities for 2016* (Jan. 11, 2016).

19 *See* Press Release, Sec. & Exch. Comm'n, *SEC Adopts Rules to Modernize Information Reported by Funds, Require Liquidity Risk Management Programs, and Permit Swing Pricing* (Oct. 13, 2016).

20 *See* Press Release, Sec. & Exch. Comm'n, *SEC Adopts Inline XBRL for Tagged Data* (Jun. 28, 2018). *See generally* Sec. & Exch. Comm'n, *Structured Disclosure at the SEC: History and Rulemaking* (last visited Dec. 3, 2022).

21 *See* Press Release, Sec. & Exch. Comm'n, *Austin Gerig Named as SEC's Chief Data Officer* (Jan. 16, 2020).

22 *See* Sec. & Exch. Comm'n, *Annual Staff Report Relating to the Use of Form PF Data* (Nov. 2, 2020).

23 *See* Sec. & Exch. Comm'n, *Staff Report on Equity and Options Market Structure Conditions in Early 2021* (Oct. 14, 2021).

24 *See* Press Release, Sec. & Exch. Comm'n, *SEC Charges Financial Services Professional and Associate in $47 Million Front-Running Scheme* (Dec. 14, 2022).

25 In this document, our use of the phrase "core data assets" refers to data assets prioritized in Data Management Program initiatives as approved by the Data Management Board.

26 In this document, we use the term "data architecture" to refer to a form of enterprise architecture comprising models, policies, rules, and standards that govern the collection, storage, arrangement, integration, and use of data in an organization such as the SEC.

27 *See*, *e.g.*, Off. of Mgmt. & Budget, Exec. Off. of the President, OMB M-22-09, *Moving the U.S. Government Toward Zero Trust Cybersecurity Principles* (Jan. 26, 2022) (calling for categorization of data based on protection needs, which would build a foundation to automate security access rules).