New York State Comptroller
THOMAS P. DiNAPOLI

Compliance With Executive Order 95

Achieving Transparency and Citizen Engagement Through Open Data

2021-D-1 | August 2021

Background

The publication of digital State data for purposes of collaboration and analysis (i.e., open data) is driven by the rationale that such practices make government more transparent and empower New Yorkers to use data to improve their lives. Open data can bring enormous benefits to the public, businesses, and researchers, and can be used in various fields such as health, food security, education, climate, intelligent transport systems, and the development of "smart cities." When open data systems are designed and implemented effectively, they allow stakeholders direct, centralized access to high-value publishable State government data to search, analyze, download, and share. This policy enables individuals, organizations, and governments to innovate and collaborate in new ways, encouraging transparency as well as public service improvements, economic value, and efficiency.

In New York, State government entities possess large amounts of valuable information on subjects such as health, business, public safety, parks and recreation, labor, and transportation. Recognizing the importance of data as a public asset, Executive Order 95 (EO 95), "Using Technology to Promote Transparency, Improve Government Performance and Enhance Citizen Engagement," issued March 11, 2013, established an Open Data website (Open Data) for the collection and public dissemination of publishable State data maintained by covered State entities (generally, those headed by individuals appointed by the Executive). Publishable data specifically includes only information that is not protected or otherwise sensitive, whose release does not violate confidentiality, privacy, or security statutes and laws. The Department of Environmental Conservation's *Recommended Fishing Rivers and Streams*, the Department of Health's *Food Service Establishment: Last Inspection*, and the Department of State's *Active Corporation*s are just a few examples of the high-value State data that EO 95 requires be published on Open Data.
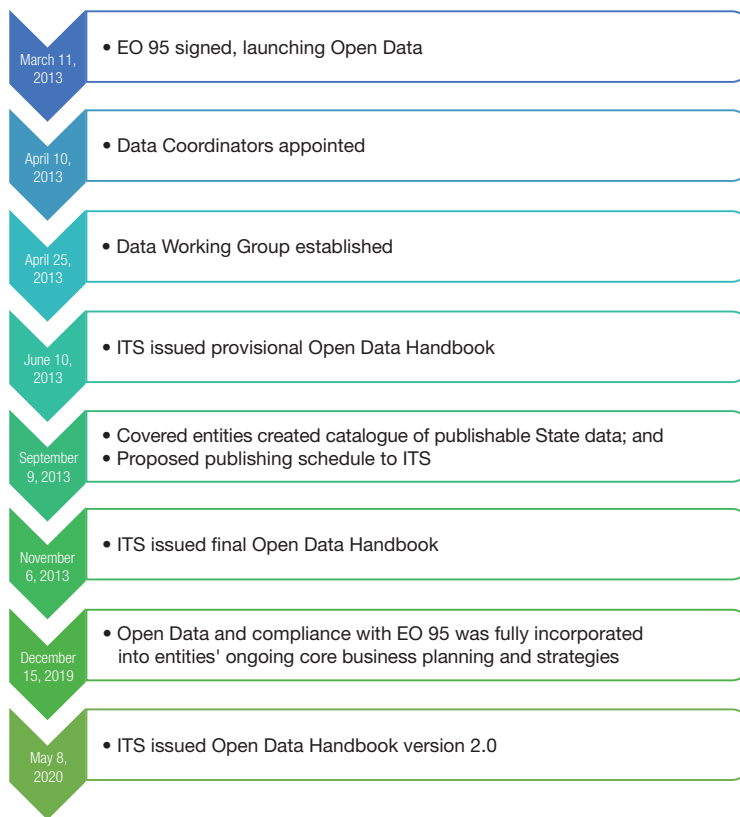
We conducted a series of five audits designed to examine State agencies' compliance with EO 95 to improve accountability and support continuous improvement of Open Data, increasing its benefits to the public and government entities.

Guidance on Open Data implementation is provided by the Open Data Handbook (Handbook). The final Handbook, issued on November 6, 2013 and updated on May 8, 2020, sets a schedule for initial and ongoing publication of data for covered State entities, and includes guidance such as a timeline for implementing EO 95 requirements, information on topics such as cataloguing and prioritizing data, and standardization and update requirements. Each covered State entity is required to create schedules and prioritize data publication in accordance with guidelines set forth in the Handbook, which directs agencies to identify high-value, high-quality, and complete data sets when creating their catalogues of data.

To coordinate EO 95 implementation, each covered State entity is required to designate a Data Coordinator. The Data Coordinator is responsible for agency compliance and serves as the liaison between the agency and the Office of Information Technology Services (ITS), the creator of the Open Data platform. ITS maintains Open Data in consultation with the Executive's Office and a Data Working Group, which is composed of representatives from ITS, the Department of State, the Division of the Budget, the Office of General Services, and up to 12 Data Coordinators representing a cross-section of covered State entities. Each covered State entity was required, within 180 days of the issuance of EO 95, to provide ITS with a catalogue of its publishable State data, along with a proposed schedule for making the data available on Open Data. The catalogue and schedule were also to be made publicly available. The Data Working Group initially met to provide guidance to agencies on implementation of EO 95, but has not met since August 2015.

EO 95 implementation was phased in beginning in March 2013, with covered entities required to be in full compliance by December 2019. The following is a high-level timeline for EO 95 implementation:

March 11, 2013	• EO 95 signed, launching Open Data
April 10, 2013	• Data Coordinators appointed
April 25, 2013	• Data Working Group established
June 10, 2013	• ITS issued provisional Open Data Handbook
September 9, 2013	• Covered entities created catalogue of publishable State data; and • Proposed publishing schedule to ITS
November 6, 2013	• ITS issued final Open Data Handbook
December 15, 2019	• Open Data and compliance with EO 95 was fully incorporated into entities' ongoing core business planning and strategies
May 8, 2020	• ITS issued Open Data Handbook version 2.0

As of May 2021, there were over 3,400 data sets on Open Data, comprising 5,200 data items such as maps, charts, and data lenses.

Overview of OSC Audits and Results

In 2019, the Office of the State Comptroller (OSC) initiated a series of audits to determine whether covered entities are complying with EO 95, including providing publishable State data to ITS under EO 95 in a timely manner, and whether data posted under EO 95 is reliable and easily usable. As of May 2021, OSC has issued five audit reports, including four reports on State agencies and one report on a public authority (collectively referred to as Agencies). OSC audited Open Data compliance at the Department of State (DOS), Office of General Services (OGS), Olympic Regional Development Authority (ORDA), Department of Environmental Conservation (DEC), and Office of Parks, Recreation and Historic Preservation (Parks). OSC determined that most Agencies were not in full compliance with EO 95, and that data was not always reliable or easily usable, limiting its value.

Compliance With EO 95

OSC's audits found that compliance with EO 95 requirements varied across the five Agencies audited. DOS generally complied with EO 95 requirements, but did not provide a complete data catalogue and schedule. OGS, Parks, and DEC took steps to meet the requirements of EO 95; however, certain aspects of EO 95 were not addressed. ORDA was not aware of and did not comply with EO 95 requirements.

Data Coordinators

Data Coordinators play a critical role in ensuring the success of the Open Data program, as they are responsible for ensuring their agency's compliance with Open Data requirements. To ensure that they have sufficient influence and expertise to carry out their responsibilities, EO 95 requires that Data Coordinators meet certain qualifications: they must have authority equal to that of a Deputy Commissioner or the head of a division or department within the entity; and they must also have knowledge of the data and resources in use by the entity.

OSC found that three of the five Agencies (DOS, OGS, and Parks) designated a qualified Data Coordinator within the required time frame. However, OGS delegated all its Data Coordinator's responsibilities and duties to an assistant. While the Handbook allows an assistant to aid the Data Coordinator in executing duties under EO 95, there was limited evidence to indicate that the assistant had the necessary knowledge of data and agency resources to meet the requirements of EO 95. DEC appointed a Data Coordinator who did not meet the required qualifications and ORDA had not appointed a Data Coordinator prior to the audit. In response to the audit, DEC and ORDA appointed Data Coordinators with the required qualifications.

Data Catalogs and Schedules

EO 95 also requires that each covered State entity create a catalogue of its publishable State data and propose a schedule to ITS for making its publishable State data publicly available. These schedules are also to be made publicly available and provide for the updating of the data catalogue as appropriate. However, OSC found that none of the five Agencies identified the total population of publishable State data that it maintained or made its schedule publicly available. ORDA was not aware of EO 95 and its requirements; while the remaining four Agencies did submit initial catalogues and schedules to ITS, the catalogues did not contain a complete population of publishable State data. During the audits, three of the Agencies (DOS, OGS, and Parks) identified and created a list of potentially publishable State data. DOS officials identified a list of 117 potentially publishable State data sets, OGS identified 35 data items, and Parks identified 69 data items. As of May 2021, DEC has not identified the total population of publishable State data it maintains, and OSC has identified 23 potentially publishable State data items on DEC's website that were not published to Open Data. DEC reviewed these data items and determined 10 of the 23 were potentially publishable and was examining them to ascertain whether they are publishable

Incorporation into Business Planning and Strategies

The Open Data Handbook required that Open Data and compliance with EO 95 be fully incorporated into State covered entities' ongoing core business planning and strategies as of December 15, 2019. OSC found that one Agency (DOS) had incorporated Open Data into its core business functions and LEAN initiatives, which are designed to improve processes and streamline government and business services. The other four Agencies (OGS, Parks, DEC, and ORDA) did not fully incorporate EO 95 into their core business functions. These Agencies did not develop policies and procedures to identify and continue publication of new high-value publishable data sets, and could not provide support that they had identified all the high-value publishable data they maintain. However, as a result of OSC's audits, Parks and DEC are both taking steps to incorporate Open Data into their ongoing business functions. Parks officials are developing procedures to conduct a biennial data survey and evaluation process to identify data sets that can be added to Open Data. DEC has created an Open Data mailbox where staff can submit data set ideas, updated data sets, and general inquiries regarding the program. Furthermore, DEC will also send out periodic reminders to staff soliciting input for Open Data suggestions and/or questions, and has added relevant information to its intranet page for staff to refer to at any time.

Data Reliability and Usability

To realize its full potential, Open Data should be accessible, delivered in a highly usable form that can be adapted to the end user's needs, and reliable. Inconsistencies and omissions reduce the data's usefulness, may lead to incorrect conclusions, and make it difficult to analyze (e.g., for performing matches or summarizing) without taking further steps. Accordingly, Open Data should provide ongoing "single-stop" access to publishable State data and make such data freely available in accessible formats for public use. Publishable State data should include comprehensive metadata and documentation to help maximize citizens' understanding, and the public should be able to easily access the data using common software applications. Further, the quality and consistency of data is essential to Open Data's value and utility. EO 95 defines quality information as appropriate, current, complete, accurate, accessible, and provided on a timely basis.

To test data reliability and usability, OSC judgmentally selected samples at each of the Agencies audited, excluding ORDA, which had not posted any data items to Open Data. The following table lists the data items OSC reviewed for usability and/or reliability during the audits (each data item is marked for the type of testing performed; the results of this review are discussed following the table):

Agency	Data Item Name	Data Reliability	Data Usability
DOS	*Active Real Estate Salespersons and Brokers*	X	X
DOS	*Active Corporations: Beginning 1800*	X	X
DOS	*Active Appearance Enhancement and Barber Business and Area Renter Licensees*	X	X
DOS	*Local Government Efficiency Program Grants: Beginning 2005*	X	X
OGS	*Master Contract Value Added Resellers Report*	X	X
OGS	*Recycled Commodities and Waste Streams: Beginning 2008*	X	X
Parks	*State Park Annual Attendance Figures by Facility, Beginning 2003*	X	X
Parks	*New York State Park Concession Contracts*	X	X
Parks	*National Register of Historic Places*	X	X
Parks	*State Park Facility Points*	X	X
Parks	*State Park Trails Map*	X	X
DEC	*Breeding Bird Atlases*		X
DEC	*Brownfield Cleanup Program, Certificates of Completion*	X	X
DEC	*Bulk Storage Facilities in New York State*	X	X
DEC	*Current Season Spring Trout Stocking*	X	X
DEC	*Environmental Remediation Sites*		X
DEC	*Harmful Algal Bloom Statewide Occurrence Summary: 2012-2018*		X
DEC	*Hudson River Estuary Program's Index of Hudson River Benthic Mapping Reports at NYS Library: 2000-2016*		X
DEC	*iMapInvasives: Invasive Plant Records*		X
DEC	*Motorized Access Trails*		X
DEC	*New York Hazardous Waste Manifest Data: 2006-2018*		X
DEC	*Office of Environmental Justice (OEJ) Grant Awards: Beginning 2006*		X
DEC	*Recommended Fishing Rivers and Streams*		X
DEC	*Solid Waste Management Facilities*	X	X
DEC	*Spill Incidents*	X	X
DEC	*Composting – Solid Waste Management Facilities*		X
DEC	*Recommended Fishing Rivers, Streams, Lakes and Ponds*		X
DEC	*Waterbody Inventory & Priority Waterbodies*	X	X

Data Reliability

Data is considered reliable if it is both accurate and complete. OSC found the data at DOS and Parks to be reliable, but found issues with the reliability of data at OGS and DEC.

Both data sets reviewed at OGS were found to have data reliability issues. For the Master Contract data, 21 of the 46 source documents reviewed (46 percent) had data that differed from what was available on Open Data. These discrepancies were caused by some description fields being entered incorrectly by OGS staff before the data was uploaded to Open Data. As a result of the audit, OGS officials stated that they have addressed this issue by implementing an internal test of the Master Contract update file prior to submission and upload to Open Data. For the Recycled Commodities data set, OGS could not provide all the documentation to support the data; therefore, OSC could not conclude it was reliable.

The audit of DEC tested six data items for completeness *(Brownfield Cleanup Program, Certificates of Completion; Current Season Spring Trout Stocking; Solid Waste Management Facilities; Bulk Storage Facilities in New York State; Spill Incidents; and Waterbody Inventory & Priority Waterbodies)* and compared the data published on Open Data to the data maintained on DEC's systems. OSC found the Spill Incidents data set had issues related to completeness. For example, 99,010 of the 287,447 facility names (34 percent) in DEC's data were not in Open Data. The audit also tested for accuracy by comparing DEC's source documents to the data posted on Open Data, and found these data sets were generally accurate, with minor exceptions (less than 10 percent).

Data Usability

The concept of data usability considers the potential for users to derive useful information from data. OSC found the data at OGS and Parks was usable, but identified usability issues with DOS and DEC data.

DOS's *Active Corporations* data set was difficult to open due to the large size of the data file. While DOS officials stated they had not received any complaints about the data set or its usability, OSC needed advanced analytical software to analyze the data, as common software could not read the complete file. DOS officials also noted that, to increase usability, they had created a data lens page for the *Active Corporations* data, which includes different visualizations, such as bar and line graphs, to facilitate understanding. However, while the data lens is valuable in helping the public visualize the data, it does not allow users to download the entire data set so it can be adapted for their specific purposes.

OSC also found each of the four data sets reviewed at DOS contained formatting inconsistencies, such as inconsistent structuring of addresses and cities. In one instance, a field contained 12 different spelling variations for the same city. In some cases, these variations occurred because DOS is required by New York Business Corporation Law to file and index corporate certificates and other instruments delivered to DOS for filing, and DOS records contain data as submitted on the certificates in its index, including any incorrect spellings. OSC also identified blank and incomplete fields.

Three of the DEC data items reviewed *(Brownfield Cleanup Program, Certificates of Completion; Current Season Spring Trout Stocking; and Spill Incidents)* had formatting issues. For example, OSC found variations in town spellings or within a date field (e.g., "March," "Mid March," "March–April"). As such, it would be difficult for the data user to aggregate information for a particular time period or town. DEC officials stated that many of these issues are due to data entry errors created when information is entered by DEC staff. They also indicated that they are investigating a fix to these issues, such as modifying the name of the data set's data column and adding drop-down boxes to help eliminate data entry errors.

Key Recommendations

In each of the Open Data audit reports, OSC made several recommendations for the Agencies to comply with all requirements of EO 95 and to improve oversight of publishable data, including the following actions when applicable:

- As soon as practicable, submit to ITS and make publicly available a complete catalogue and schedule of publishable State data for addition to Open Data.

- Work with ITS to enhance the quality and utility of publishable State data by improving data accessibility, as well as consistency, where permitted under applicable law.

- Develop procedures to:

  - Ensure data is reliable and usable.

  - Incorporate compliance with EO 95 into the Agency's core business practices.

  - Ensure accurate, complete, and timely updates to data posted on Open Data.

In their responses, DOS and OGS disagreed with the recommendations and stated they believed they had fully complied with EO 95. ORDA, DEC, and Parks agreed with the recommendations and provided information regarding steps they have taken to implement the recommendations.

Conclusion

While OSC found that four of the five Agencies audited were complying with EO 95 provisions or taking steps toward full compliance, greater sharing of valuable data in usable formats by all Agencies would promote transparency and innovation. Based on the intent of EO 95 – to "enhance public access to government data and make government in New York State more transparent" – agencies should make every effort to make as much high-value publishable State data as practicable publicly available.

Contact

Office of the New York State Comptroller
110 State Street
Albany, New York 12236

(518) 474-4044

www.osc.state.ny.us

Prepared by the Division of State Government Accountability

Executive Team:
Tina Kim, Deputy Comptroller
Ken Shulman, Assistant Comptroller

Audit Team:
Brian Reilly, Audit Director
Heather Pratt, Audit Manager
Theresa Podagrosi, Audit Manager
Amanda Dare, Examiner-in-Charge
Mary McCoy, Supervising Editor
Andrea Majot, Senior Editor

f y ⊙

Like us on **Facebook** at facebook.com/nyscomptroller
Follow us on **Twitter** @nyscomptroller
Follow us on **Instagram** @nys.comptroller