

# Central Data Exchange System Identity Data Are Unreliable

March 5, 2024 | Report No. 24-N-0025



## Report Contributors

Tertia Allen  
Yoon An  
LaSharn Barnes  
Troy Givens  
Nii-Lantei Lamptey  
lantha Maness  
Christina Nelson  
Teresa Richardson  
Scott Sammons  
Michelle Wicker

## Abbreviations

CDX	Central Data Exchange
EPA	U.S. Environmental Protection Agency
NIST	National Institute of Standards and Technology
OIG	Office of Inspector General
OMS	Office of Mission Support
RMAM	Registration Maintenance Account Manager

## Cover Image

The Central Data Exchange is the EPA's electronic reporting site for environmental data. (EPA OIG adaptation of EPA images of mountains, leaves, and abstract binary code)

**Are you aware of fraud, waste, or abuse in an EPA program?**

**EPA Inspector General Hotline**

1200 Pennsylvania Avenue, NW (2431T)  
Washington, D.C. 20460  
(888) 546-8740  
(202) 566-2599 (fax)  
[OIG.Hotline@epa.gov](mailto:OIG.Hotline@epa.gov)

Learn more about our [OIG Hotline](#).

**EPA Office of Inspector General**

1200 Pennsylvania Avenue, NW (2410T)  
Washington, D.C. 20460  
(202) 566-2391  
[www.epaoig.gov](http://www.epaoig.gov)

Subscribe to our [Email Updates](#).  
Follow us on X (formerly Twitter) [@EPAoig](#).  
Send us your [Project Suggestions](#).



**OFFICE OF INSPECTOR GENERAL**  
U.S. ENVIRONMENTAL PROTECTION AGENCY

March 5, 2024

**MEMORANDUM**

**SUBJECT:** Central Data Exchange System Identity Data Are Unreliable  
Report No. 24-N-0025

**FROM:** Sean W. O'Donnell, Inspector General *Sean W O'Donnell*

**TO:** Kimberly Patrick, Principal Deputy Assistant Administrator  
Office of Mission Support

Jennie Campbell, Director  
Office of Information Management  
Office of Mission Support

The U.S. Environmental Protection Agency Office of Inspector General [initiated](#) an audit to review the EPA's Central Data Exchange, or CDX, access security controls. While conducting work on that audit, which remains ongoing, we identified issues with data quality and data integrity that may negatively affect the EPA's decision-making and communication of programmatic information about the environment. We decided to issue this management alert to inform the EPA of the issues we identified because they could impact the Agency's ability to fulfill its mission and carry out its regulatory obligations.

**This management alert supports the following EPA mission-related effort:**

- Operating efficiently and effectively.

**This management alert addresses the following top EPA [management challenge](#):**

- Managing grants, contracts, and data systems.

You are not required to respond to this management alert because it contains no recommendations. If you submit a response, however, it will be posted on the OIG's website, along with our memorandum commenting on your response. Your response should be provided as an Adobe PDF file that complies with the accessibility requirements of section 508 of the Rehabilitation Act of 1973, as amended. The final response should not contain data that you do not want to be released to the public; if your response contains such data, you should identify the data for redaction or removal along with corresponding justification.

We will post this report to our website at [www.epaoig.gov](http://www.epaoig.gov).

*To report potential fraud, waste, abuse, misconduct, or mismanagement, contact the OIG Hotline at (888) 546-8740 or [OIG.Hotline@epa.gov](mailto:OIG.Hotline@epa.gov).*

## Background

The EPA has over 30 disparate information systems that record and store environmental data.<sup>1</sup> To streamline and facilitate data reporting, the EPA developed a central web-based registration and reporting system, called the CDX, that allows companies, states, tribes, and other regulated entities to electronically register their identity and, if applicable, exchange their environmental data with the Agency from a single place. Currently, the CDX accepts environmental data for the EPA's air, water, hazardous waste, and toxics release inventory programs, which then can be sent to one or more of the other EPA information systems connected to the CDX.

The environmental data submitted to the CDX comply with the requirements of the environmental laws that govern the EPA's regulatory responsibilities, such as the Safe Drinking Water Act and the Resource Conservation and Recovery Act. For example, data in the EPA's Safe Drinking Water Information System, governed by the Safe Drinking Water Act, identify violations of drinking water regulations by public drinking water systems, and data in the EPA's Resource Conservation and Recovery Act Information system, governed by the Resource Conservation and Recovery Act, track the retrieval, transportation, and disposal of hazardous waste.

Reporting data in compliance with the applicable environmental laws begins with the regulated entity creating a CDX account and requesting access to the EPA's environmental systems. To register and create a CDX account, the regulated entity provides identity data, such as an individual or entity name, a physical address, an email address, and a phone number. An EPA employee or contractor serving as the registration maintenance account manager, or RMAM, uses the identity data to grant the regulated entity access to the environmental systems. After being granted access, a regulated entity is referred to as a CDX user and can post environmental data to the CDX. The CDX transfers the identity data to the EPA's 30-plus environmental systems and the environmental data to the applicable systems, as needed, to support specific programs. RMAMs use the CDX user identity data to contact and further verify the identity of the CDX user, if required by EPA environmental regulations.

The EPA follows Agency policy and federal guidance to ensure adherence to data quality and integrity requirements. EPA Directive No. 2150-P-17.2, *Information Security—Interim System and Information Integrity Procedures*, dated January 17, 2017,<sup>2</sup> required that system personnel in the EPA Office of Mission Support's Information Exchange Services Branch shall, among other steps, "verify the checks for input validation as part of system testing" and "configure the information system to check all arguments or input data strings submitted by users." And according to the National Institute of Standards and Technology's *NIST Big Data Interoperability Framework: Volume 7, Standards Roadmap*, NIST Special Publication 1500-7r2, dated October 2019, cleaning data is the "keystone for data quality" and is necessary to provide accurate analytic outputs. Further, the *NIST Framework* indicates that data are

---

<sup>1</sup> According to the CDX registration webpage, a user can elect to register for an estimated 37 primary program services or systems.

<sup>2</sup> EPA Directive No. 2150-P-17.2 was the applicable Agency procedure in effect for the scope of this project. It was superseded by EPA Directive No. 2150-P-17.3 on November 21, 2023.

clean when they are free from inconsistencies when errors, such as incorrect data types, have been addressed.

## Responsible Office

The EPA Office of Mission Support owns the CDX system. The Information Exchange Services Branch within the Office of Mission Support is responsible for operating and managing the CDX to lead the Agency in its electronic data exchange and to support the Agency's mission to protect human health and the environment.

## Scope and Methodology

We conducted our work from August 2022 to August 2023. While our overall audit, which is still ongoing, is being conducted in accordance with generally accepted government auditing standards, the work related to this management alert does not constitute an audit done in accordance with these standards. However, we did follow the OIG's quality control procedures to ensure that the information in this report is accurate and supported.

The OIG Office of Investigations alerted the OIG Office of Audit to CDX account identity issues, which led to our findings regarding the CDX's identity data in the RMAM and CDX user files. We also reviewed a March 2023 report, *EPA Data Challenges and Opportunities*, which was issued by the EPA's Data Governance Council and identified challenges with the Agency's data governance. We then analyzed the identity data contained in the RMAM and CDX user files. At the time of our analysis, the RMAM file contained 1,873 records that included RMAM communication data, such as email addresses and phone numbers. The CDX user file contained 195,950 records that included CDX user identity data, such as first and last names, physical addresses, and organization names. To identify and analyze the data integrity issues identified within this report, we obtained assistance from the OIG Data Analytics Directorate and reviewed only the files that contained identity data. It is possible that other CDX files have similar data quality issues.

## OIG Concerns

We identified instances of unreliable data in the RMAM and CDX user files that met neither the EPA quality and integrity requirements in effect during our project scope nor the NIST quality and integrity guidance. Data quality and integrity are interrelated and ensure data accuracy, completeness, validity, consistency, and fitness for purpose. The Agency can minimize the issues that we identified by putting measures in place within the CDX to check the identity data entered by CDX users.

### *CDX User Data Appear to Be Unvalidated and Unreliable*

The RMAM and CDX user files contained data that did not meet the quality and integrity requirements outlined in the *NIST Framework* guidance and EPA Directive No. 2150-P-17.2. In other words, the user-

submitted data did not always appear to have been validated for accuracy or quality. For example, in the CDX user file, we found users with questionable first and last names, such as “<i>YOU'REACKED</i>” and “abcdefghijklmn.” We requested the account creation date, account status, and account history for several of the CDX user accounts to determine if suspicious activity occurred within the CDX system. The Agency did not respond with sufficient information to verify the account status and account activity for these questionable accounts. In the RMAM file, we identified phone numbers listed as “1231231233.” Table 1 illustrates the types of issues we noted in the files. We may identify additional issues as we complete our audit.

**Table 1: Examples of issues identified in the RMAM and CDX user files**

Examples	File and field	Issues	Number of issues identified	Percent of records with issues (%)
abc@123.com gmail.com yahoo.com	RMAM file* Email address	Questionable email addresses. Some emails appear to be personal email addresses used by EPA personnel or those conducting business on the EPA’s behalf; however, EPA guidance strongly discourages the use of personal emails.	122	6.51
5555555555 1231231233 9999999999	RMAM file* Phone number	Questionable phone numbers. The data appears to be false because the phone numbers have the same sequence of numbers.	280	14.94
<i>YOU'REACKED</i> Aaaaaa aatest<>	CDX user file† First name	Questionable first names. One name reads “YOU'REACKED”. First names rarely have repetitive letters and symbols.	94	0.05
<b>YOU'REACKED</b> abcdefghijklmn aa123<>	CDX user file† Last name	Questionable last names. Last names rarely have sequenced letters of the alphabet, numbers, or symbols.	79	0.04
CDX Testing Company Test_23 <>marquee 1.00E+11	CDX user file† Organization name	Questionable organization names with symbols and other noncharacters.	71	0.04
Numbers like 1,2,7,10 Firstname.lastname@163.com xcvzv Xenias 4 3/15/2001	CDX user file† Physical address	Questionable addresses, with entries including personal email addresses instead of physical addresses. Some entries contained numbers with no street names or random characters.	599	0.31

Source: OIG analysis of EPA CDX data.

\* The RMAM file that we reviewed contained 1,873 records.

† The CDX user file that we reviewed contained 195,950 records.

### ***Unreliable System Data May Affect EPA Decision-Making***

The issues that we identified within the RMAM and CDX user files, along with the Office of Investigations’ observations regarding CDX identity issues, could indicate the presence of fraudulent accounts in the CDX. For example, someone may have hacked into the CDX and changed the data or used a legitimate user’s data to create a fraudulent account. The Agency was unable to provide supporting documentation to verify whether the questionable data that we identified were not associated with potential fraudulent accounts that the Office of Investigations identified.

Although the quantity of issues indicated in Table 1 may not appear significant, it takes only one instance of fraud to negatively affect information systems. Threat actors could potentially use fraudulent accounts to gain access to not only the CDX but also the EPA’s environmental systems connected to the CDX. The data from the environmental systems support the EPA’s program services, and the credibility of the environmental data that are submitted and aggregated affects the EPA’s programs’ ability to support the Agency’s strategic plan.

The *NIST Framework* guidance states that not having clean data can lead to inaccurate analytics, incorrect conclusions, and wrong decisions. While we reviewed only the files that contained identity data, it is possible that other CDX files have similar data quality issues. CDX data are transferred across the EPA’s environmental systems and subsequently used by the EPA to make decisions and advance its strategic plan goals. If the EPA does not mitigate its CDX data integrity issues, it cannot provide assurance that its environmental data are accurate and reliable.

## **Agency Response and OIG Assessment**

On October 31, 2023, the EPA Office of Mission Support responded to our draft report, generally disagreeing with our findings. The Agency’s response is included in Attachment A, and we detail our assessment of this response below.

We maintain our conclusion that the EPA is not in compliance with Agency policy and federal guidance that govern data quality and integrity. The Agency stated that an “entry in the users table” means only that an entity created a user account and had a valid email address.<sup>3</sup> It further stated that it is “not appropriate to assume that the quality of data in the user file has any relationship with the quality of any other data traversing through CDX.” However, this response does not address the issues that we identified in the report regarding data quality and integrity in the RMAM and CDX user files. Although the Agency stated in its response that the CDX and its “connected systems have extensive business requirements to ensure data meet the data quality requirements that are specific to each programmatic data flow,” we did not see evidence that CDX data quality controls were operating as intended, as evidenced by the RMAM and CDX user file issues identified in Table 1 of this report.<sup>4</sup>

The EPA needs to ensure that the data residing in the CDX are clean and free from inconsistencies and errors, as advised by the *NIST Framework* guidance. The Agency’s response attributed some of the data quality and integrity issues to “tester data” and acknowledged that such test data must be better categorized so it is not confused with official data. However, when we provided the EPA an opportunity to address the questionable data, it did not provide supporting documentation to verify the presence of test data. If the EPA does not or cannot distinguish its test data, which appear questionable, from the official data, it will continue to have an inaccurate representation of data in its RMAM and CDX user files.

---

<sup>3</sup> See Attachment A, “OMS Response to Report Concerns” table, No. 4.

<sup>4</sup> See Attachment A, “OMS Response to Report Concerns” table, No. 4. We will review data quality and integrity control documentation during our overall audit.

We maintain that the data in the RMAM and CDX user files did not always appear to have been validated for accuracy or quality. The Agency said that it validates email addresses and further stated that it requires identity proofing for the “submitter” role to make sure that users cannot submit data to the CDX “without a properly initialized or valid account;” however, we concluded that email validation does not verify the identity of a person. Further, an Agency representative stated that identity proofing for all submitters was not conducted.<sup>5</sup>

We also maintain that the issues we identified could indicate the presence of fraudulent accounts in the CDX. The Agency stated that our examples in Table 1 are of “raw data” and do not consider context or business processes.<sup>6</sup> It also stated that “[d]ata in the user file are not indicative of other data that are processed via CDX.”<sup>7</sup> However, the Office of Mission Support did not provide supporting documentation to show that the CDX data included raw or test data, nor did the office provide evidence of compensating controls that mitigate the risks of storing questionable data in the CDX. The appearance of questionable data in the CDX user fields may indicate that the EPA’s CDX data integrity and quality controls or rules are not successfully preventing questionable data.

The Agency also raised concerns regarding specific terminology and information included in our draft report, and we addressed these concerns and updated our report as appropriate.

cc: Michael S. Regan, Administrator

Janet McCabe, Deputy Administrator

Dan Utech, Chief of Staff, Office of the Administrator

Wesley J. Carpenter, Deputy Chief of Staff for Management, Office of the Administrator

Faisal Amin, Agency Follow-Up Official (the CFO)

Andrew LeBlanc, Agency Follow-Up Coordinator

Susan Perkins, Agency Follow-Up Coordinator

Jeffrey Prieto, General Counsel

Tim Del Monico, Associate Administrator for Congressional and Intergovernmental Relations

Nick Conger, Associate Administrator for Public Affairs

Vaughn Noga, Chief Information Officer and Deputy Assistant Administrator for Information Technology and Information Management, Office of Mission Support

Helena Wooden-Aguilar, Deputy Assistant Administrator for Workforce Solutions and Inclusive Excellence, Office of Mission Support

Dan Coogan, Deputy Assistant Administrator for Infrastructure and Extramural Resources, Office of Mission Support

Stefan Martiyan, Director, Office of Continuous Improvement, Office of the Chief Financial Officer

---

<sup>5</sup> See Attachment A, “OMS Response to Report Concerns” table, No. 5 and 7.

<sup>6</sup> See Attachment A, “OMS Response to Report Concerns” table, No. 6.

<sup>7</sup> See Attachment A, “OMS Response to Report Concerns” table, No. 6.



Yulia Kalikhman, Acting Director, Office of Resources and Business Operations, Office of Mission Support

Tonya Manning, Director and Chief Information Security Officer, Office of Information Security and Privacy, Office of Mission Support

Shari Grossarth, Office of Policy OIG Liaison

Stuart Miles-McLean, Office of Policy GAO Liaison

Michael Benton, Audit Follow-Up Coordinator, Office of the Administrator

Afreeka Wilson, Audit Follow-Up Coordinator, Office of Mission Support

## Agency Response to Draft Report



### OFFICE OF MISSION SUPPORT

WASHINGTON, D.C. 20460

October 31, 2023

#### MEMORANDUM

**SUBJECT:** Management Response to Office of Inspector General Draft Report “Central Data Exchange System Identity Data Are Unreliable” Project No. OA-FY23-0095 dated October 16, 2023

**FROM:** Vaughn Noga, Chief Information Officer, NOGA  
Deputy Assistant Administrator for Information Technology and Information Management

**TO:** LaSharn Barnes, Director  
Information Resources Management Directorate  
Office of Audit, Office of Inspector

VAUGHN

Digitally signed by  
VAUGHN NOGA  
Date: 2023.10.31  
14:36:06 -04'00'

Thank you for the opportunity to respond to the concerns in the October 16th draft report titled “Central Data Exchange System Identity Data Are Unreliable” that outlined the OIG’s concerns. Specifically, that:

1. “CDX user data appear to be unvalidated and unreliable;” and
2. “Unreliable system data may affect EPA decision-making”

Following is a summary of the agency’s position on each of the concerns outlined in the report as well as additional context and information regarding the Central Data Exchange’s (CDX) data validation procedures, including validations for user information. Generally, we disagree with the assertions and conclusions that data in the user file in any way are indicative of other data that are processed via CDX. There are data rules both within CDX and the respective program systems to ensure data quality is maintained and those additional controls are not considered in this evaluation and report.

We do appreciate the investigation the IG team is performing as it highlights the need for CDX to better categorize “tester” data so that it is not confused with official data.

OMS RESPONSE TO REPORT CONCERNS

No.	OIG Concern or Statement	OMS' Response
1.	“The EPA has 49 disparate environmental systems that record and store environmental data.”	Do not concur. OIM is unsure how the OIG calculated “49 disparate environmental systems.”
2.	“data in the EPA’s Safe Drinking Water Information System, also governed by the Clean Water Act”	Typo - Data in the Safe Drinking Water Information System is governed by the Safe Drinking Water Act, not the Clean Water Act.
3.	“The Information Exchange Solutions Branch within the Office of Mission Support”	Typo – should be the Information Exchange Services Branch within the Office of Mission Support
4.	“We identified several instances of unreliable data that did not meet quality and integrity requirements in the RMAM files and CDX user file data we reviewed. CDX data issues affect not only the CDX but also any systems that capture data from the CDX. In other words, any questionable data in the CDX could be shared with the EPA’s 49 environmental systems that are connected to the CDX”	Do not concur. An entry in the users table only means that an individual has created a user account and that they had a valid email address. It does not mean that they have access to or can provide data to any of the EPA systems. RMAMs are responsible for ensuring only authorized users have access to these systems. It is also not appropriate to assume that the quality of data in the user file has any relationship with the quality of any other data traversing through CDX. CDX as well as the connected systems have extensive business requirements to ensure data meet the data quality requirements that are specific to each programmatic data flow.
5.	“In contrast to the NIST Framework and EPA Directive No. 2150-P-17.2, many RMAM files and CDX user files contained data that did not meet data quality and integrity requirements. In other words, the user-submitted data did not always appear to have been validated for accuracy or quality.”	Do not concur. The current CDX account creation process requires email validation, which sends an email to the provided email address before the account can be initialized and used. Depending on the role, identity proofing may also be required.
6.	“In our professional opinion, the issues that we identified within the RMAM and CDX user files along with the complaints from the Office of Investigation regarding CDX fraudulent identity issues, could indicate the presence of fraudulent accounts in the CDX.”	Do not concur. The issues identified are of the raw data and without considerations of the context or business processes.  The data examples provided in Table 1 are of raw data and without context. When other data elements that make up an RMAM or CDX user are evaluated more completely, it is obvious the issues raised are without merit. For example: 1. The domain 123.com is valid and

		<p>there is no standing to claim it as erroneous or fake.</p> <p>2. For EPA staff or contractors testing the system as an industry user – they are testing explicitly not as an EPA employee. When EPA based business is being tested EPA email is used, which complies with guidance.</p> <p>Most of the users associated with other data examples are test accounts or are inactive.</p>
7.	<p>“Threat actors could use these accounts to gain access to not only the CDX, but the 49 EPA environmental systems connected to the CDX. CDX data integrity is critical to ensure that only legitimate users have access to the CDX.”</p>	<p>Do not concur. CDX has additional controls in place to protect against this. For example, CDX does have an email validation process which protects CDX from having invalid email addresses on valid accounts. For the “submitter” role, identity proofing is required, therefore users are not able to submit data to CDX or any of the connected systems without a properly initialized or valid account.</p>
8.	<p>“CDX data integrity issues also pose a risk to the quality of the EPA’s decision-making and communications. The NIST Framework states that not having clean data can lead to inaccurate analytics, incorrect conclusions, and wrong decisions. While we reviewed only the files that contained the identity data, it is possible that other CDX files have similar data quality issues. CDX data are transferred across EPA’s environmental systems and used by the EPA, regulated entities, and the public use to make programmatic decisions and meet regulatory requirements. If the EPA does not mitigate its CDX data integrity issues, it cannot provide assurance that its environmental data are accurate and reliable.”</p>	<p>Do not concur. Data in the user file are not indicative of other data that are processed via CDX. There are data rules both within CDX and the respective program systems to ensure data quality is maintained and those additional controls are not considered in the evaluation and report.</p>

If you have any questions regarding this response, please contact Marilyn Armstrong, Audit Follow-up Coordinator, of the Office of Resources and Business Operations, (202) 564-1876 or [armstrong.marilyn@epa.gov](mailto:armstrong.marilyn@epa.gov).

Cc: Tertia Allen  
Yoon An  
LaSharn Barnes  
Troy Givens  
Nii-Lantei Lamptey  
Iantha Maness  
Christina Nelson  
Teresa Richardson  
Scott Sammons  
Michelle Wicker  
Erin Collard  
Austin Henderson  
David Alvarado  
Jennie Campbell  
Dwane Young  
Joe Carioti  
Dan Coogan  
Yulia Kalikhman  
Marilyn Armstrong



## Whistleblower Protection

U.S. Environmental Protection Agency

*The whistleblower protection coordinator's role is to educate Agency employees about prohibitions against retaliation for protected disclosures and the rights and remedies against retaliation. For more information, please visit the OIG's whistleblower protection [webpage](#).*

### Contact us:



**Congressional Inquiries:** [OIG.CongressionalAffairs@epa.gov](mailto:OIG.CongressionalAffairs@epa.gov)



**Media Inquiries:** [OIG.PublicAffairs@epa.gov](mailto:OIG.PublicAffairs@epa.gov)



**EPA OIG Hotline:** [OIG.Hotline@epa.gov](mailto:OIG.Hotline@epa.gov)



**Web:** [epaoig.gov](http://epaoig.gov)

### Follow us:



**Twitter:** [@epaoig](https://twitter.com/epaoig)



**LinkedIn:** [linkedin.com/company/epa-oig](https://linkedin.com/company/epa-oig)



**YouTube:** [youtube.com/epaoig](https://youtube.com/epaoig)



**Instagram:** [@epa.ig.on.ig](https://www.instagram.com/epa.ig.on.ig)



[www.epaoig.gov](http://www.epaoig.gov)