Privacy Impact Assessment
for
Collibra Data Intelligence Cloud (Collibra)

December 2024

Point of Contact
Kenneth Janoski
Division of Data (DOD)/
Data Operations and Standards Branch

Reviewing Official
Kellie Cosgrove Riley
Chief Privacy Officer

OVERVIEW

The Collibra Data Intelligence Cloud ("Collibra") is a Software-as-a-Service (SaaS) product that will enable the Commodity Futures Trading Commission ("CFTC" or "Commission") to build an Enterprise Data Catalog and manage the metadata for mission-critical data sets. Collibra will connect to various data sources and ingest the metadata, which will be used to build data lineage and facilitate data classification activities. The data stewards for each respective date set will be able to use Collibra to add and maintain technical and business glossary terms and to build workflows to enforce data policies.

The CFTC's Division of Data ("DOD") is responsible for Collibra. Members of other Divisions across the Commission may use Collibra, but, other than vendor technical support personnel responding to a specific CFTC request for assistance, no persons external to CFTC will have access to Collibra or the information therein. Use of Collibra will facilitate the execution of CFTC's statutory mission by enhancing the CFTC's understanding and, therefore, its use of its existing and future data sets. This privacy impact assessment is being conducted because Collibra collects metadata from various data sets that the Commission receives from its regulated entities and those data sets from which it extracts metadata may contain personally identifiable information (PII).

The mission data sets from which Collibra is being used to extract metadata are located in the CFTC's Order Book on Demand – Market Financial Data in the Cloud authorization boundary ("OBOD-MFDC") and the SQL server in the CFTC's Azure General Support System (GSS) boundary. Collibra uses an Edge server, i.e., a server that is within the Collibra authorization boundary but is physically located in OBOD-MFDC in proximity to the relevant data sets, and business intelligence tools to accomplish the metadata extraction. The Edge server uses Java database connectivity, an application programming interface, to connect to the data sets and pull in a portion of the data set needed to extract the metadata. Once collected, the data is encrypted within the Edge server and cannot be accessed by anyone within the CFTC or at the vendor.

The Collibra Edge server collects data from the data sets in OBOD-MFDC and Azure GSS and the Collibra business intelligence tools extract the metadata. The extracted metadata is then pushed to the Collibra cloud environment, outside of CFTC's network authorization boundary, and is accessible to all CFTC employees and contractors via the CFTC Data Hub. The original data pulled into the Edge server is not pushed to the Collibra cloud and, instead, remains on the Edge server for 24-48 hours at which time is it purged.

CFTC users are granted role-based access to Collibra. All CFTC employees and contractors have read-only access to the metadata and the resulting data catalog in Collibra via the CFTC Data Hub. Certain users, primarily within DOD and Information Technology, have additional administrative access to, for example, configure the Edge server and the business intelligence tools to facilitate the metadata harvesting, develop dashboards, and otherwise monitor and administer Collibra. Data Stewards and Data Analysts across the CFTC's mission divisions will be trained and given access to Collibra beyond read-only on an as-needed basis. All users, regardless of their role-based permissions, may access Collibra only using CFTC issued laptops and credentials.

1. AUTHORITIES AND OTHER REQUIREMENTS

1.1 What legal authority(ies) permit or require the CFTC to collect, use, maintain, or disseminate the information in the system, project, or program?

The collection of the information contained in the OBOD-MFDC data sets that Collibra accesses to extract meta data, as the extraction and collection of that meta data, is done in pursuit of the Commission's mission and consistent with its jurisdiction as set forth in the Commodity Exchange Act, 7 U.S.C. §§ 1, et seq.

1.2 Is there a Privacy Act System of Records Notice(s) (SORN(s)) that applies to the information in the system, project, or program? If yes, please identify it. If no, please explain why not.

The various data sets that Collibra accesses in OBOD-MFDC may contain Privacy Act-covered records subject to various CFTC SORNs, to include CFTC-10, Investigatory Records, and CFTC-15, Enterprise Surveillance, Oversight & Risk Monitoring. However, the information in Collibra, to include that pulled into the Edge server and the metadata pushed to the Collibra cloud has no applicable SORN. Information about an individual collected by the Edge server is not, and cannot, be searched or retrieved by an individual's name or other personal identifier. Moreover, the metadata extracted from the data sets in the Edge server and pushed to the Collibra cloud is not information about an individual and is not subject to the Privacy Act.

1.3 Please identify the applicable accreditation boundary evaluated during the Authority to Operate (ATO) process. If the system, or involved project or program, is hosted by a Cloud Service Provider (CSP), please identify the CSP and state whether the system has

Federal Risk and Authorization Management Program (FedRAMP) authorization.

Collibra is FedRAMP certified, and the CFTC issued an Authorization to Operate (ATO) for Collibra on January 23, 2024.

1.4 Is there a records retention schedule(s) approved by the National Archives and Records Administration (NARA) that governs the retention of the records maintained in the system, project, or program? If yes, please identify the schedule. If not, please explain.

Yes. The records that are pulled into the Edge server for metadata extraction are governed by NARA General Records Schedule (GRS) 5.2, Transitory and Intermediary Records, Item 010, Transitory Records, and are destroyed when no longer needed for business use or according to an agency predetermined period or business rule. The extracted metadata that is pushed to the Collibra cloud is governed by GRS 3.1.051, Data Administration Records, and are destroyed five years after a project is completed or superseded, with longer retention permitted if required for business use.

1.5 Please list all forms or other information collection instruments that are used to collect the information in the system. If the information collected is subject to the Paperwork Reduction Act, provide the name of the form or information collection and the applicable OMB control number.

Collibra does not collect any information directly from the public and is not subject to the Paperwork Reduction Act.

2. INFORMATION

2.1 Identify the information that the system, project, or program collects, maintains, uses, or disseminates.

The information that Collibra collects via the Edge server may vary among the data sets in OBOD-MFDC and Azure GSS from which it is deployed to extract metadata. It will potentially include business contact information from individual traders or their points of contact, such as name, phone number, email address, and mailing address, and identifying numbers, such as Legal Entity Identifiers and National Futures Association IDs.

2.2 Is the information collected directly from the individual and, if not, what are the sources of information collected. How is the information obtained?

The information in Collibra is not collected directly from the individual. Rather, it is collected from data sets in OBOD-MFDC and Azure GSS via the Edge server using Java database connectivity, an application programming interface.

2.3 What steps are taken to minimize the collection of information in the system, project, or program?

The Collibra Edge server collects some but not all of the information in the data sets in OBOD-MFDC and Azure GSS. The tool is designed and configured to pull in only a portion of the data set in order to facilitate the extraction of the metadata. While no explicit effort is made to further minimize the information pulled into the Edge server, because to do so would potentially limit the amount and/or type of metadata to be extracted, that information is deleted within 24-48 hours of collection.

2.4 Discuss how the accuracy, integrity, and/or quality of the information identified above are determined.

Collibra is not designed to confirm the accuracy, integrity, or quality of the data it extracts from the OBOD-MFDC and Azure GSS data sets; it is only designed to identify and extract the meta data. Rather, Collibra relies on the original data set source of data for the data's accuracy, integrity, and quality.

2.5 Privacy Impact Analysis: Related to the Collection and Characterization of the Information

Generally, the risks associated with the collection and characterization of information concern minimization, i.e., collecting only that information needed to achieve the business function, and the accuracy, quality, or integrity of the information collected.

Collibra via the Edge server collects a sample of a data set in order to extract metadata; although the Edge server is configured to pull in a subset of a data set, no explicit effort is made to further minimize the information pulled into the Edge server because to do so would potentially limit the amount and/or type of metadata to be extracted. Any risk that more information than is needed will be pulled into the Edge server, however, is mitigated by adherence to a short retention period pursuant to which the information is deleted from the server within 24-48 hours of collection.

The accuracy, quality, and integrity of the data is not addressed within Collibra. Collibra collects data only for the purpose of extracting the associated metadata and the accuracy, integrity, or quality of the data collected via the Edge server is presumed as it is pulled from the authoritative source.

The privacy risks associated with collection and characterization of CFTC data is low.

3. USE OF THE INFORMATION

3.1 Describe how the information is used internally at the CFTC to support a business purpose.

The information pulled into the Edge server from the data sets in OBOD-MFDC and Azure GSS is used to extract meta data that describes the content of the data sets. The metadata, in turn, is used to facilitate the execution of CFTC's statutory mission by enhancing its understanding and, therefore, use of its existing data sets. The extracted metadata will be used to build an enterprise data catalog and to facilitate improved data quality and data governance across the Commission.

3.2 Identify all programs or offices, and an approximate number of individuals, within the CFTC that have assigned roles and responsibilities with respect to the system, project, or program or with whom information is shared. Explain how each uses the information.

All CFTC employees and contractors have read-only access to the metadata and the resulting data catalog in Collibra via the CFTC Data Hub. Certain users, primarily DOD and Information Technology (IT) staff, have additional administrative access to configure the Edge server and the business intelligence tools to facilitate the metadata harvesting, develop dashboards, address any technical issues and assist others across the CFTC with using Collibra. Data stewards and data analysts across the CFTC's mission divisions will be trained and given access to Collibra beyond read-only on an as-needed basis.

3.3 Privacy Impact Analysis: Related to the Internal Access and Use

Generally, privacy risks associated with internal access and use concern the access and use of information by an unauthorized person or the use of the information by an authorized user for an unauthorized purpose. These risks

are mitigated here via the use of role-based access controls; in fact, most users of the information collected and used by Collibra have read-only access to only the meta data in the CFTC Data Hub and no access to the Edge server. Those data stewards and data analysts who are granted access to the Edge server may apply for access to use Collibra on a data set to which they already have a need-to-know and for the sole purpose of extracting the associated metadata. Moreover, the data pulled into the Edge server is encrypted and no user has any visibility into the contents of the Edge server. The CFTC administrator for Collibra and the system owner control access to the tool and review the access roles annually to ensure all continue to have an official need to access/use Collibra. Accordingly, the privacy risk associated with internal access and use of Collibra and the data it collects is low.

4. INFORMATION SHARING

4.1 Is the information shared outside of the CFTC as part of usual agency operations? If so, identify the organizations or individuals with whom the information is shared, how and *why* it is shared, and how it is to be used.

None of the information, including but not limited to any PII, collected by the Collibra Edge server is shared outside of the CFTC; even within CFTC, no user has visibility into the content of the Edge server. The metadata, which does not include any PII, is pushed to the Collibra cloud (the CFTC Data Hub) where CFTC users have read only access to it.

4.2 How is the information sharing described in 4.1 compatible with the purpose for which it was originally collected described in 3.1?

No PII collected by Collibra is shared outside of the CFTC.

4.3 Is the information sharing documented in a Memorandum of Understanding, other information sharing agreement, or a contract?

No PII collected by Collibra is shared outside of the CFTC.

4.4 Are there limitations on re-dissemination of the information by the third party in place, either in the documents identified in 4.2 or otherwise?

No PII collected by Collibra is shared outside of the CFTC.

4.5 Describe how the system, project, or program accounts for and maintains a record of disclosures outside of the CFTC.

No PII collected by Collibra is shared outside of the CFTC.

4.6 Privacy Impact Analysis: Related to Information Sharing

No PII will be shared outside of CFTC from Collibra, and, therefore, this is no privacy risk associated with external sharing of PII.

5. DATA RETENTION

5.1 Explain how long and for what reason the information is retained.

In accordance with NARA GRS 5.2, which permits transitory records to be retained for as long as they are needed for a business purpose, the data collected by the Edge server is retained for 24-48 hours in order to extract metadata.

5.2 Describe how the information is disposed of at the end of its retention period.

All records are purged from the Collibra Edge server on a rolling schedule every 24-48 hours.

5.3 Privacy Impact Analysis: Related to Retention

Generally, the privacy risks related to retention concern retaining information for longer than is necessary to achieve the business purpose for the collection. Here that risk is mitigated, and, therefore, low because the information pulled into the Edge server is retained for just 24-48 hours in order to meet the business purpose of extracting the metadata.

6. NOTICE

6.1 How are individuals provided notice prior to the collection of their information? If they are not provided notice, please explain why.

Individuals whose information may be collected via the Edge server do not interact with or have access to Collibra and, therefore, Collibra does not provide any notice regarding the collection. Notice regarding Collibra and the information it collects is provided through the publication of this PIA.

6.2 Do individuals have the opportunity to consent to the use of their information or decline to provide their information? Please explain.

Collibra does not collect information directly from individuals and the individuals whose information is collected in Collibra's Edge server do not have the opportunity to consent to the collection.

6.3 Privacy Impact Analysis: Related to Notice

Generally, the privacy risk related to notice concerns whether the individual was aware of and had the opportunity to consent, either explicitly or implicitly, to the collection and use of their information. Here, individuals do not interact directly with Collibra and no direct notice or opportunity to consent is provided. However, this risk has little to no privacy impact on the individual given that the information is used only to extract metadata and not to make any decisions about or provide any service or benefit to the individual.

7. REDRESS

7.1 What procedures are in place that allow individuals to access their information?

Individuals whose information may be collected by the Collibra Edge server do not have access to that information and no procedure exists to provide such access.

7.2 What procedures are in place to allow individuals to correct/amend inaccurate or irrelevant information about themselves?

Individuals whose information may be collected by the Collibra Edge server do not have access to that information and no procedure exists for individuals to request access or to request that information in Collibra be corrected or amended.

7.3 How are individuals notified about the procedures for correcting/amending their information?

Individuals whose information may be collected by the Collibra Edge server do not have access to that information and no procedure exists for individuals to request access or to request that information in Collibra be corrected or amended.

7.4 Privacy Impact Analysis: Related to Redress

Generally, the privacy risks related to redress concern an individual's ability to participate in the collection and use of their information, e.g., by accessing their information and being able to request amendment or correction of that information where necessary. Here, individuals do not interact directly with Collibra and the business purpose for which Collibra may have their information does not pertain to the individual. There is, therefore, little to no privacy impact on the individual given that the information is used only to extract metadata, is deleted in short order, and is not used to make any decisions about or provide any service or benefit to the individual that may be impacted by incorrect information.

8. AUDITING AND ACCOUNTABILITY

8.1 How is the information in the system secured?

Physical, administrative, and technical safeguards are in use to secure Collibra. Physical safeguards include physical access procedures such as onsite guards and use of ID badges. Administrative safeguards include limiting "write" privileges to those approved and licensed by the authorized administrative user to only those users with a demonstrated need, yearly audits to ensure all licensed users still need access to the system, contingency planning, and requiring users to acknowledge and sign CFTC Rules of Behavior (ROB). Technical safeguards include enabling access to Collibra only from a CFTC issued laptop and through CFTC's virtual private network (VPN), requiring multifactor authentication (MFA) using a CFTC-issued personal identity verification (PIV) card, firewall, and encryption for data in transit and at rest.

8.2 Are audit logs generated reflecting access and use of the information in the system? If not, please explain. If so, please describe how often they are reviewed and for what purpose.

Yes, audit logs are for the Collibra Edge server, a Linux server, are generated and include the following:

- Date/Time changes;
- User/Group Account Modifications;
- Network information Modifications (hostname, system locale);
- Access Control Modifications;
- Session Initiation Events;
- File Access – Failure;
- Successful File System Mounts;
- File Deletion Events; and

- Kernel Module Loading and Unloading.

Because the only use of CFTC data in Collibra is to extract the associated metadata, event logs do not indicate how Collibra data is used.

These logs are sent in real time to Sentinel and CFTC's Enterprise Security Operations Center (eSOC) reviews Sentinel log data related to alerts received by eSOC and related to alerts triggered by the source from which data in the logs are gathered.

8.3. What procedures are in place to determine who has access to the system, and are they documented?

All CFTC staff have read-only access to the metadata in the Collibra cloud (CFTC Data Hub) when using a CFTC-issued device accessed via the CFTC VPN with their PIV card. A Community Manager within DOD is responsible for evaluating and approving access beyond read-only for those individuals who demonstrate a need and can satisfy the relevant criteria.

8.4 Describe what privacy training is provided to users, either general or specific to this system or project.

All CFTC employees and contractors are required to take annual security and privacy awareness training, the completion of which is a prerequisite to a user gaining and maintaining access to the CFTC network. In addition, all CFTC employees and contractors must sign the CFTC Information Technology Rules of Behavior Acknowledgment annually.

8.5 Privacy Impact Analysis: Related to Auditing and Accountability

As outlined above, audit logs for Collibra are generated and are monitored by eSOC. DOD is accountable for providing appropriate system access to those individuals who need elevated permissions beyond read-only. Therefore, the privacy impact related to auditing and accountability is low.