U.S. Consumer Product Safety Commission PRIVACY IMPACT ASSESSMENT	
Name of Application/System:	CPSC Data Lake
Office/Directorate:	EXIT
Date:	1/27/2021
A. CONTACT INFORMATION	
Person completing PIA: (Name, title, organization and ext.)	Ryan Powell, Systems Architect, Sapient
System Owner: (Name, title, organization and ext.)	Padma Chari, Director - Solutions Development, EXIT, x7018
System Manager/Technical POC: (Name, title, organization and ext.)	Jason Jen, IT Specialist - Systems Architect, EXIT, x7724

B. APPROVING OFFICIALS

Padma Chari Digitally signed by Padma Chari Date: 2021.04.30 14:13:45 -04'00'

System Owner — Date

Senteria Jones Digitally signed by Senteria Jones Date: 2021.04.30 14:18:43 -04'00'

Privacy Officer — Date

PATRICK MANLEY Digitally signed by PATRICK MANLEY Date: 2021.05.03 09:47:44 -04'00'

Chief Information Security Officer (CISO) — Date

Abioye Mosheim Digitally signed by Abioye Mosheim Date: 2021.04.29 16:03:06 -04'00'

Asst. General Counsel for FOIA, Records and Privacy — Date

JAMES ROLFES Digitally signed by JAMES ROLFES Date: 2021.05.03 12:03:52 -04'00'

Senior Agency Official for Privacy (SAOP) — Date

C. SYSTEM OF RECORDS NOTICE

1. Will the system or application maintain records that contain information about individuals? Yes ☑ No ☐

2. Will the system or application allow records to be retrieved by a personal identifier*? Yes ☑ No ☐

If the answer to both 1 and 2 above is "Yes," then the system requires a System of Records Notice (SORN).

* *A personal identifier might include an individual's name, address, email address, telephone number, social security number, photograph, biometric information, or any other unique identifier that can be linked to an individual.*

D. PRIVACY ASSESSMENT

1. Will the system or application be used to collect, store, or transmit personally identifiable information (PII)*?

Yes ☑ No ☐ (If there is **NO** information collected, stored, or transmitted that is identifiable to the individual, the remainder of this PIA does not have to be completed.)

* *Personally Identifiable Information (PII) as defined in OMB Memorandum M-07-16 refers to information that can be used to distinguish or trace an individual's identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. (e.g. Social Security numbers, Passport numbers, etc.)*	
2. Generally describe the type of information that will be collected, stored, or transmitted.	The CPSC Data Lake is a system that allows access to a copy of data from other systems. The Data Lake does not collect information.
3. What categories of individuals are covered in the system? (public, employees, contractors)	The CPSC Data Lake may contain information related to the public, employees, and contractors. Access to the Data Lake will be controlled via application roles related to each data set in the Data Lake. The system will be accessible to CPSC employees and contractors with the appropriate roles.
E. SYSTEM DATA	
4. Is the PII collected verified for accuracy? Why or why not?	Verification of data accuracy is expected to be a function of the system of record. The Data Lake may perform additional verifications as part of the data ingest pipeline, but the Data Lake will not perform updates to the data. Since data ingest pipelines will run regularly, new or changed data in the source systems will be reflected in the Data Lake.
5. Is the PII current? How is this determined?	Currency of the data is expected to be handled by the system of record. The Data Lake may implement data ingest controls based on the currency of data.
6. Who will be responsible for protecting the privacy of the individuals whose PII is collected, maintained, or shared on the system? Have policies and/or procedures been established for this responsibility and accountability?	The CPSC Data Lake is responsible for protecting all information contained within, using security controls of equal or higher level to the system of record. Maintenance of data will be the responsibility of the source system owner. Storage of data is the responsibility of the team operating the Data Lake. Policies and procedures for accessing the data will be established and approved by the Data Management Working Group (DMWG), and access will be restricted by default.
7. Is there a process for individuals to have inaccurate PII that is maintained by the system corrected or amended, as appropriate?	The process to correct or update data would be to engage the process for the source system. The CPSC Data Lake does not correct or ammend data, rather, that data is brought over directly from source systems.
8. Is the source of the information from the individual or is it taken from another source? If not directly from individual, then what other source?	The source of all information will be from another source, namely another CPSC system. The CPSC Data Lake is designed to aggregate data across all of CPSC's data systems.
F. MAINTENANCE AND ADMINISTRATIVE CONTROLS	

9. What opportunities do individuals have to decline to provide information or to consent to particular uses of the information?	The opportunity to decline to provide information is the same as in the system of record. The CPSC Data Lake will only ingest data that exists in another CPSC system.
10. Do other systems that interconnect to the system share, transmit, or access the PII in the system? If yes, explain the purpose for system to system transmission, access, or sharing of PII.	The CPSC Data Lake will allow users to access and transmit PII through data queries. The purpose of the CPSC Data Lake is to provide a copy of production data for sharing and reporting across the business on an as-needed basis. Access to data will be provided and revoked by the Data Owner for each data source (other CPSC system).
11. What involvement will contractors have with the design and maintenance of the system? Has a contractor confidentiality agreement or a Non-Disclosure Agreement (NDA) been developed for contractors who work on the system?	The CPSC Data Lake system will be designed, developed, and maintained by the contractor Publicis Sapient in conjunction with EXIT. A Non-Disclosure Agreement has been established and is enforced for all contractors working on the CPSC Data Lake.
12. What are the retention periods of PII for this system? Under what guidelines are the retention periods determined? Who establishes the retention guidelines?	PII retention periods are the same as their system of record. The CPSC Data Lake will only maintain a copy of data provided by other CPSC systems. If data is added or removed from the system of record, the CPSC Data Lake will be updated to reflect the change.
13. What are the procedures for disposition of the PII at the end of the retention period? How long will any reports that contain PII be maintained? How is the information disposed (e.g., shredding, degaussing, overwriting, etc.)?	Reports containing PII will be refreshed based on the latest version of data stored within the CPSC Data Lake. Once data is removed from the system of record, it will be permanently removed from the CPSC Data Lake. Backups of data within the CPSC Data Lake will be purged after 30 days. Information is syncronized with production source systems using the Azure Data Factory.
14. Is this system currently identified as a CPSC system of records? If so, under which notice does the system operate?	The CPSC Data Lake is not a system or record. The Data Lake maintains a copy of data from systems of record only. What will be moved into the Data Lake is all of data from the systems like IFS, ITDS/RAM, CPSRMS. There will be a data pipeline setup that connects the Data Lake to each of the on prem databases so that the data in the Data Lake is kept up to date. Data in the Data Lake will be primarily used for analysis,
G. ACCESS TO DATA	
15. Who will have access to the data in the system? (e.g., contractors, managers, system administrators, developers, other).	Data Access to system data will be granted by the Data Owner for each system of record that the Data Lake stores information for. Access could include both CPSC employees and CPSC contractors. In addition, any Azure cloud administrators operating the production environment would have access to the system. Development will happen in a different environment, separate from production data.

16. What controls are in place to prevent unauthorized access to the data?	Role-based access controls, and the use of two-factor authentication based on the PIV card will prevent unauthorized access. Roles are validated for every query/report of data contained within the CPSC Data Lake.
17. What controls are in place to prevent the misuse (e.g., browsing) of PII by those having access?	Data access will be controlled by the data owners of data. Each data owner may restrict access to their data to prevent misuse.
18. Is access to the PII being monitored, tracked, or recorded?	Access to all data in the CPSC Data Lake will happen via cloud-based API calls that are monitored, tracked, and recorded.
19. For CPSC support staff, how is access to the PII determined? Are criteria, procedures, controls, and responsibilities regarding access documented? Does access to PII require manager approval?	Access to PII requires approval of the data owner for each associated system of records. The criteria, procedures, controls, and responsibilities documented for each system of record will be standardized for the CPSC Data Lake through policies developed in conjunction with the Data Management Working Group.
20. What third-party organizations will have access to the PII (if known)? Who establishes the criteria for what PII can be shared?	All access to the CPSC Data Lake will require a CPSC Azure account, PIV card, or Data Lake API credentials. All connections to the CPSC Data Lake require authorization by the Data Owner for each system of record.
21. What CPSC personnel roles will have access to PII fields (e.g., users, managers, system administrators, developers, contractors, other)?	Users who have been granted specific roles by the Data Owner for each system of record will be able to access PII for that specific system of record.
22. Will any of the personally identifiable information be accessed remotely or physically removed?	Access to information in the CPSC Data Lake will happen via an Express Route (Equinix) connection to the Azure cloud. Access of information from other locations is not planned. There will be no physical movement of data for the CPSC Data Lake.