



PII CATALOG PRIMER: 4 Steps to Building a PII Catalog for Privacy

A PII catalog gives organizations a single source of truth of what, where, and whose personal data they're storing, processing and analyzing.

Table of contents

Introduction	3
Inventory	5
Correlate	6
Classify	7
Map	8
Conclusion	9

Introduction

Data catalogs rely on ingestion of metadata from databases to help companies organize, describe and, well, catalog their data – from tables to files to schemas.

Metadata is data about data, and can be aggregated and interpreted to provide a range of valuable business input. It's often integrated with data dictionaries and glossaries for business initiatives, or with data ownership metadata for data governance.

Metadata alone, however, doesn't tell the whole story: it's removed from the data values themselves, varies significantly by platforms, and can't identify who that data belongs to.



For organizations to better manage data, metadata catalogs should be supplemented with a PII catalog.



For organizations to more easily manage, monitor, and protect personal data, metadata catalogs should be supplemented with a PII catalog.

A PII catalog provides a foundational layer for understanding how data is associated with individuals and entities – and leverages metadata for added context. PII catalogs enable organizations to implement data security best practices, maintain continuous compliance, and automate policy to help manage and protect their data.

In order to build a PII catalog, organizations need a catalog of customer data – a one-stop-shop for data inventory. A PII catalog gives organizations a single source of truth of what, where, and whose personal data they're storing, processing and analyzing.



A PII catalog gives organizations a single source of truth of what, where, and whose personal data they're storing, processing and analyzing.



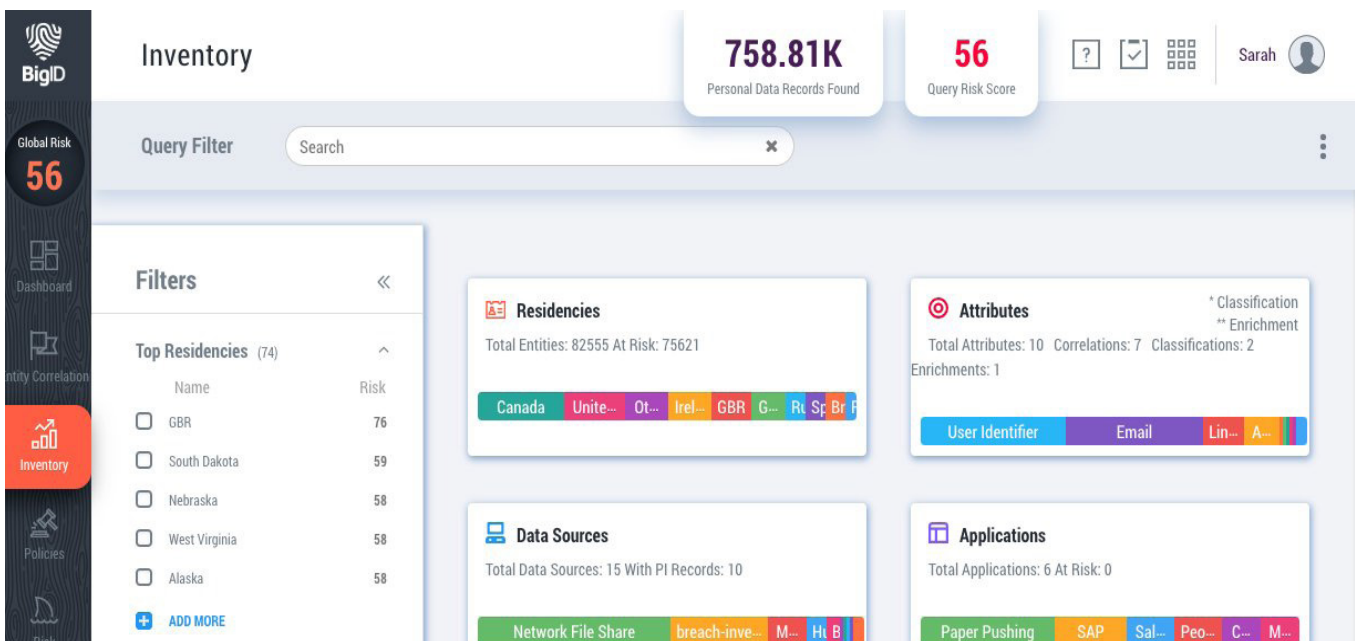
Step 1: Inventory Your Data

A decentralized data inventory is the first step to a PII catalog: organizations need to be able to generate a dynamic and easily navigable inventory of personal data without duplicating, moving, copying, or compromising that data.

A data inventory should display:

- where personal data records are located,
- their attributes and categorization,
- residency,
- the entity or individual with whom they are associated,
- and related metadata.

BigID’s privacy-centric data discovery empowers organizations to find, correlate and classify any personal data – from names to credit card numbers to favorite sandwich shop – and layer that identity intelligence into a PII catalog.



The screenshot shows the BigID Inventory dashboard. At the top, it displays 'Inventory' with a search bar and a 'Query Filter' dropdown. Key metrics are shown: 758.81K Personal Data Records Found and 56 Query Risk Score. The user 'Sarah' is logged in. The dashboard is divided into several sections:

- Filters:** A sidebar on the left showing 'Top Residencies (74)' with a list of countries and their risk scores: GBR (76), South Dakota (59), Nebraska (58), West Virginia (58), and Alaska (58). There is an 'ADD MORE' button at the bottom.
- Residencies:** A card showing 'Total Entities: 82555 At Risk: 75621' with a horizontal bar chart for various countries including Canada, United States, etc.
- Attributes:** A card showing 'Total Attributes: 10 Correlations: 7 Classifications: 2 Enrichments: 1' with a horizontal bar chart for attributes like User Identifier, Email, etc.
- Data Sources:** A card showing 'Total Data Sources: 15 With PI Records: 10' with a horizontal bar chart for sources like Network File Share, etc.
- Applications:** A card showing 'Total Applications: 6 At Risk: 0' with a horizontal bar chart for applications like Paper Pushing, SAP, etc.

Once a personal data inventory is established, organizations can drill down into PI and PII findings, retrieve identity and entity records from across all data sources, and easily establish an audit trail.

In turn, analysts can assign labels and tags to the attributes that upstream services can consume to further manage and protect that data. Or, in the case of catalogs that allow for data element crowdsourcing, those business terms can be consumed, propagated and validated against an accurate accounting based on data values.



BigID discovers dark data and correlates it to an identity, surfacing identity intelligence for data relationships across an organization's data stores.



Step 2: Correlate Your Data

Privacy requirements (whether as an ethical constraint or as a regulatory compliance mandate) introduce the need to understand data elements in the context of who the data belongs to, along with the need to determine whether data is personal information based on association with an individual.

BigID discovers dark data and correlates it to an identity, surfacing identity intelligence for data relationships across an organization's data stores.

By correlating personal data, organizations will get deeper insight into their PII catalog, and associate personal information into an identity or entity profile.

Step 3:

Classify Your Data

Personal data isn't limited to personally identifiable information (PII), but extends to personal information (PI). Data classification, therefore, needs to go beyond regular expressions and structured data patterns, and extend to classifying personal data of all types.

The ability to automatically and accurately classify data at scale across data sources for both business and technical audiences lies at the heart of why enterprises invest in tools like data catalogs.

BigID's approach is to create a uniform model for how metadata is stored and organized, and apply context-specific classification so that organizations can get more value from their data. Once a data registry is compiled from data values that incorporates correlation to establish entity association, applying metadata insights can significantly improve the accuracy of classification.

BigID leverages advanced machine learning to automatically find, map, and classify any type of data across any data store, giving organizations full visibility into their data.

Traditional classification tools are optimized for specific data stores, often focusing on either structured or unstructured data. BigID, on the other hand, scans a wide range of data stores from unstructured data to big data to NoSQL to SaaS and more, for a true cross-platform holistic view.

Step 4: Map Your Data

Data maps are a practical requirement for data privacy impact assessments (PIA), and are a critical component of many data protection regulations. They map business flows across an organization – showing how data flows across entities and business users throughout a data lifecycle.

BigID helps build out a data map of personal data – with contextual data intelligence rather than surveys – so that organizations can more easily analyze and assess risk on how personal information is collected and processed.

This PII mapping serves as the foundational point of reference for tools catering to business users, who can then add further notation or data curation, and identify where compliance issues can emerge.

By tying the PIAs to maps that are the output of ongoing scans, organizations shorten the time involved with updating PIAs, while taking a proactive stance to privacy compliance.



What You Can Do with a PII Catalog

Just as enterprises look to machine learning to extract insights from their data, that same approach can provide data intelligence by delivering insights into relationships between datasets and how an organization is making use of data. Machine learning can be leveraged to inventory, catalog, and populate a comprehensive data registry, all while automating classification via metadata analysis.

Once an organization has accurately established a PII catalog – and identified, inventoried, mapped, and classified all personal data across their data stores – they’re able to automate policy, implement data security best practices, and maintain continuous compliance.

Want to see how BigID can help build your organization’s personal data inventory?

Get started at [BigID.com/demo](https://www.bigid.com/demo).