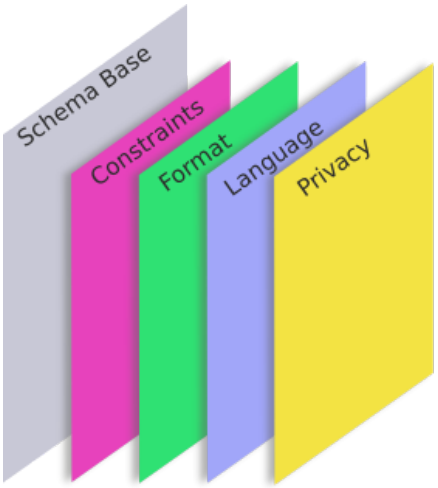
Cloud Privacy Labs

Layered Schema Architecture for Semantic Processing

Layered Schema Architecture is developed to help machines understand the underlying meaning of information by embedding semantic annotations into data. It provides a mechanism to define unified data semantics as a **schema base** and to add additional **layers** to represent variations for different data sources, contexts, and jurisdictions. These embedded semantics offer significant advantages to traditional systems where the meaning is “hardwired” into the processing logic. With traditional systems, processing logic must be modified when two systems need to interoperate in a new way or new information needs to be exchanged. With **Layered Schema Architecture**, integrating new systems or changing data models is achieved by simply adding schema variations.

Traditional schemas describe the structure of data with limited semantic information. Layered schema architecture slices a traditional schema into multiple **layers** with each layer adding both structural and semantic information. The **schema base** is the first layer that defines the structure and the vocabulary used to capture and process data. The data model can be flat for tabular, linked with other schemas, or nested to represent more elaborate structures such as health data. Schema **layers** add semantic annotations to the base model. These annotations include:



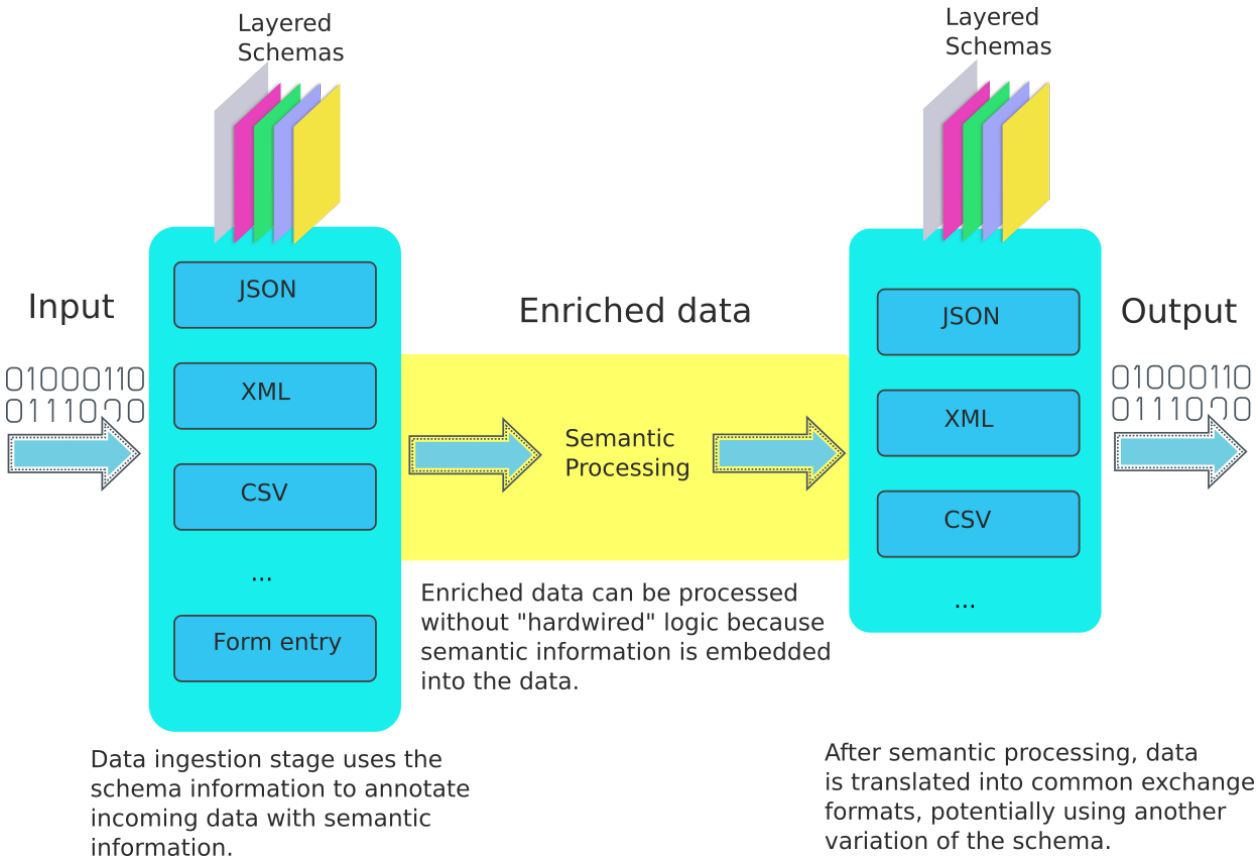
- Constraints: Required attributes, length limits,...
- Format: Phone number with area code, date/time,...
- Language: English, Spanish,...
- Privacy classifications: Personally identifiable information, sensitive information,...
- Retention policies: Attributes must be cleared after a period,...
- Provenance: Data source, signatures, ...

Layered schemas improve interoperability between different applications by enabling semantic processing of data coming from heterogeneous sources. Each data source uses a variation of the schema to interpret data based on a common vocabulary. Privacy and security layers mark attributes for encryption, de-identification, and consent validation during data exchange. Data provenance layers add metadata to keep track of data source and ownership. For data warehousing and data pooling operations, layered schema architecture unifies the vocabulary for the ingested data and adds the necessary semantic annotations to decouple meaning from the processing logic.

info@cloudprivacylabs.com

Cloud Privacy Labs

Layered schemas enable auto-generated data entry forms to capture data using a common vocabulary. These forms can be built in multiple languages and with different data entry conventions. Since captured data will include the semantic annotations of the schema layers, data processing applications do not have to deal with hardwired data semantics.



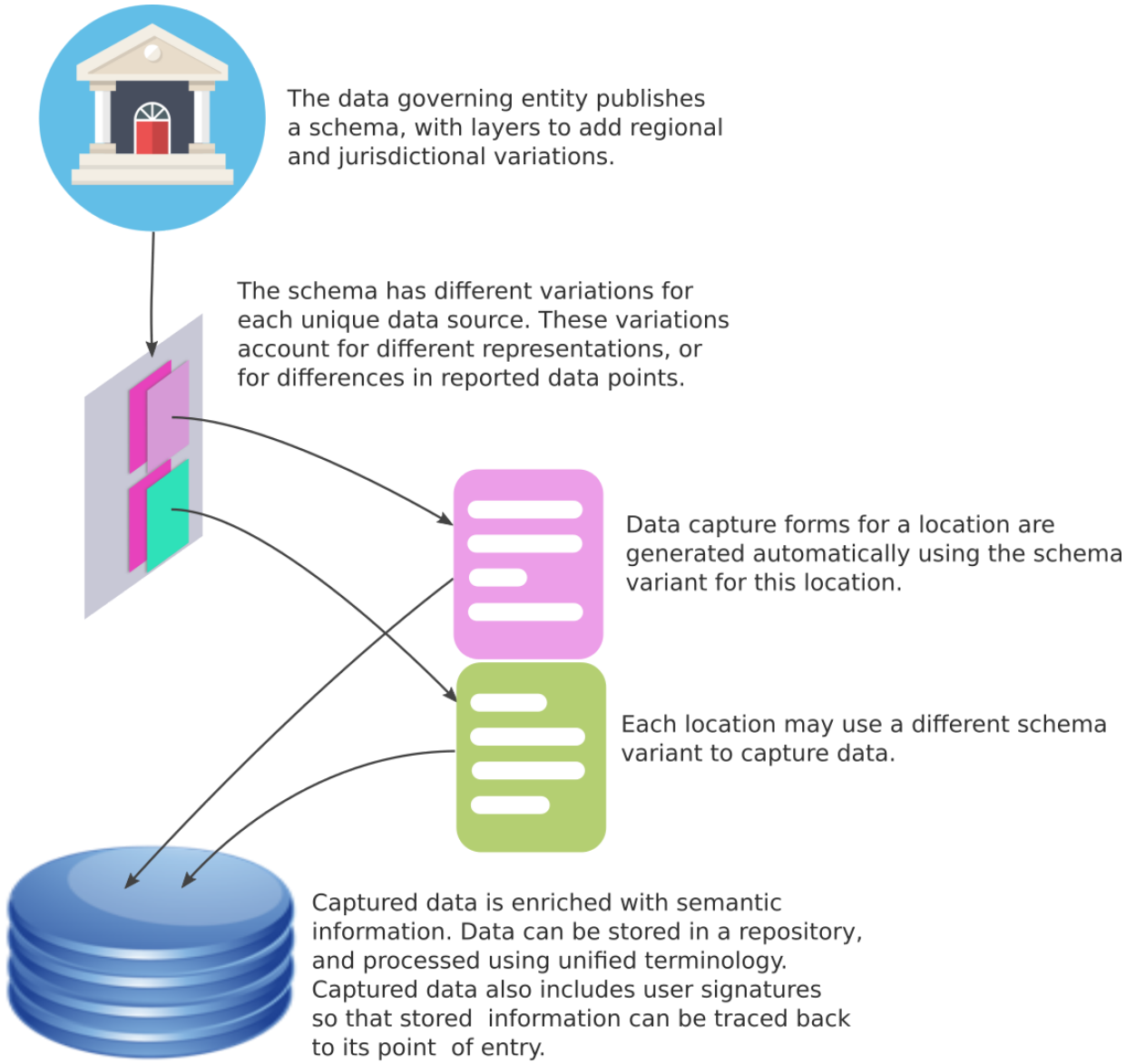
Semantic processing of data from heterogeneous sources with layered schemas

info@cloudprivacylabs.com

Cloud Privacy Labs

Use Case: Data Collection with Data Entry

This use-case of layered schema architecture uses auto-generated data capture forms to collect data from different sources. The applications for this use-case include but are not limited to data reporting with manual entry, data collection, research subject recruitment, and questionnaires. This collection process accounts for variations among data collection points without additional effort for customization while still attaining a unified data vocabulary as a result.

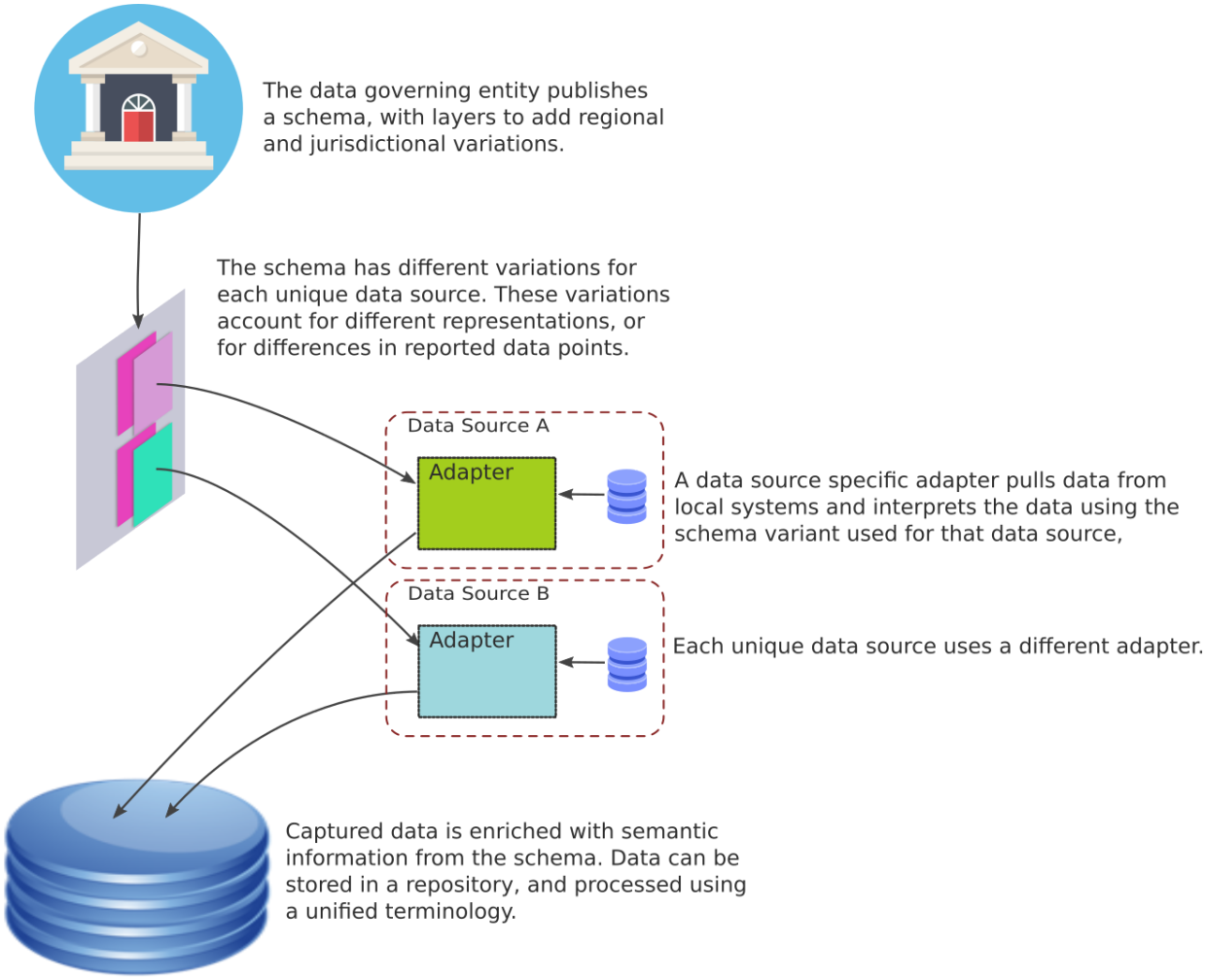


info@cloudprivacylabs.com

Cloud Privacy Labs

Use Case: Data Collection with API Integration

This use-case of layered schemas relies on **adapters** to collect data from heterogeneous sources using API or database integrations. The applications for this use-case include data consolidation, data warehousing, information exchange hubs, and health information exchange. Data captured from multiple sources can be unified using a common vocabulary and enriched with semantic annotations.



info@cloudprivacylabs.com

Cloud Privacy Labs

Use Case: Data Exchange

In this use case, data exchange between organizations is governed by agreements that specify what data items will be exchanged, whether or not certain data elements such as personally identifiable information or sensitive data can be exchanged, and if so under what conditions.