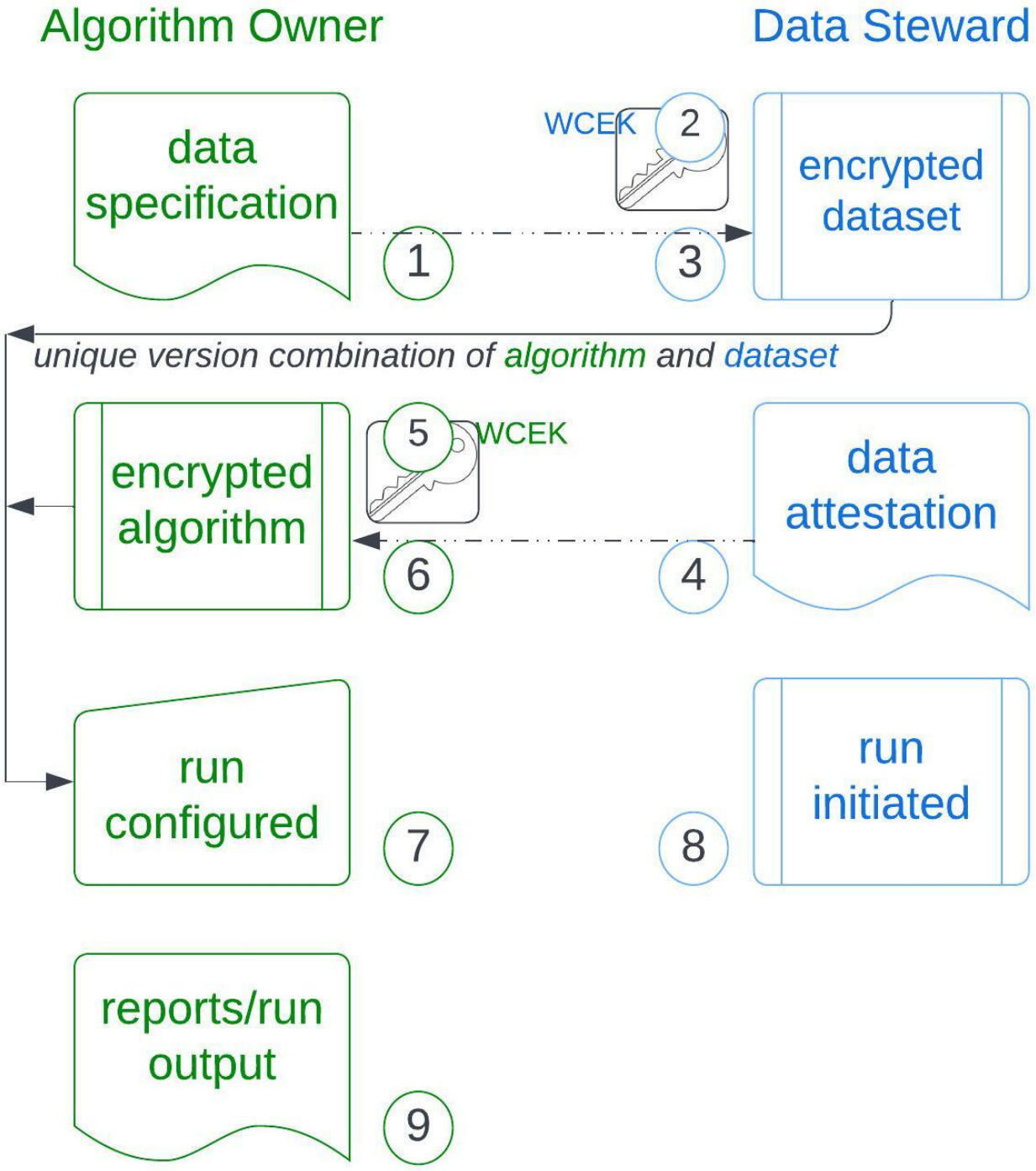
EscrowAI Flow

alex.lange

Exported on 11/21/2023

Table of Contents

EscrowAI enables two-party collaboration between a Data Steward and an Algorithm Owner for processing AI/ML algorithms on sensitive data.



1 EscrowAI collaborative workflow

The following describes the steps between the two collaborating parties. (Dotted arrows in the diagram show relationships among project artifacts.)

1. The Algorithm Owner uploads a *data specification* artifact, which is a text document that describes the data they require to process their algorithm.

2. The Data Steward receives the data specification document and curates the data based on the data specification. After curation, the Data Steward encrypts the dataset using their own Content Encryption Key (CEK) and uploads the dataset to blob storage. At this time the Data Steward also wraps the CEK to create a Wrapped Content Encryption Key (WCEK) and uploads that key to EscrowAI.

   EscrowAI supports datasets of various types, namely:
   - Structured
   - Unstructured
   - Image
   - Video/Audio
   - Omics data

3. A Shared Access Signature (SAS) URL pointing to the the encrypted data in blob storage is configured in EscrowAI by the Data Steward.
4. After curating the dataset the Data Steward uploads a *data attestation* artifact, which is a text document that describes the curated dataset. The dataset is thus associated with its corresponding data specification and the data attestation Artifacts.
5. After viewing the data attestation artifact, the Algorithm Owner can modify their algorithm and encrypt it using their own CEK and wrap the CEK to create a WCEK, which is then uploaded to EscrowAI.
6. The Algorithm owner uploads the encrypted algorithm. At this step, the Algorithm Owner must select the algorithm type they are uploading. Algorithm type determines the input, output, and workflow process in EscrowAI.

7. When all artifacts needed to process the algorithm on the dataset have been uploaded, a run configuration is created associating the Algorithm version and the dataset version. A run configuration is a unique combination of Algorithm and dataset version. No two run configurations can have the same algorithm and dataset version.
8. A run is an instantiation of the run configuration and deployed with end-to-end encryption on Confidential Computing nodes in the Data Steward's environment. Multiple Runs of a particular run configuration can be deployed but not in parallel.