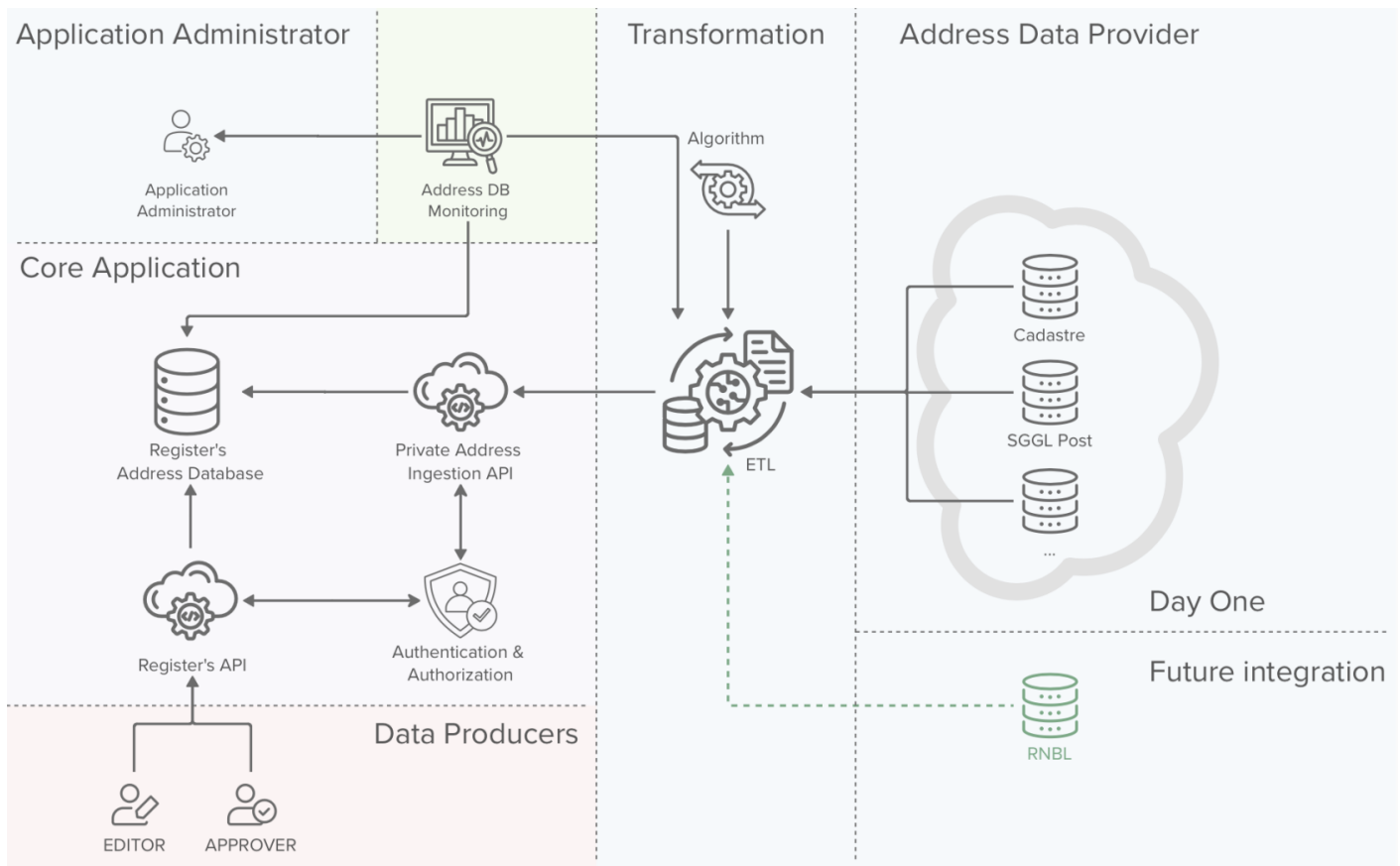


7.4. Address ingestion via the ETL process



The background colors of the above image are to be interpreted as:

- **purple:** the core application a.k.a. backend that will be built in the context of this project
- **green:** supporting systems that will be used in the context of this project
- **blue:** trusted parties
- **red:** external users of the system

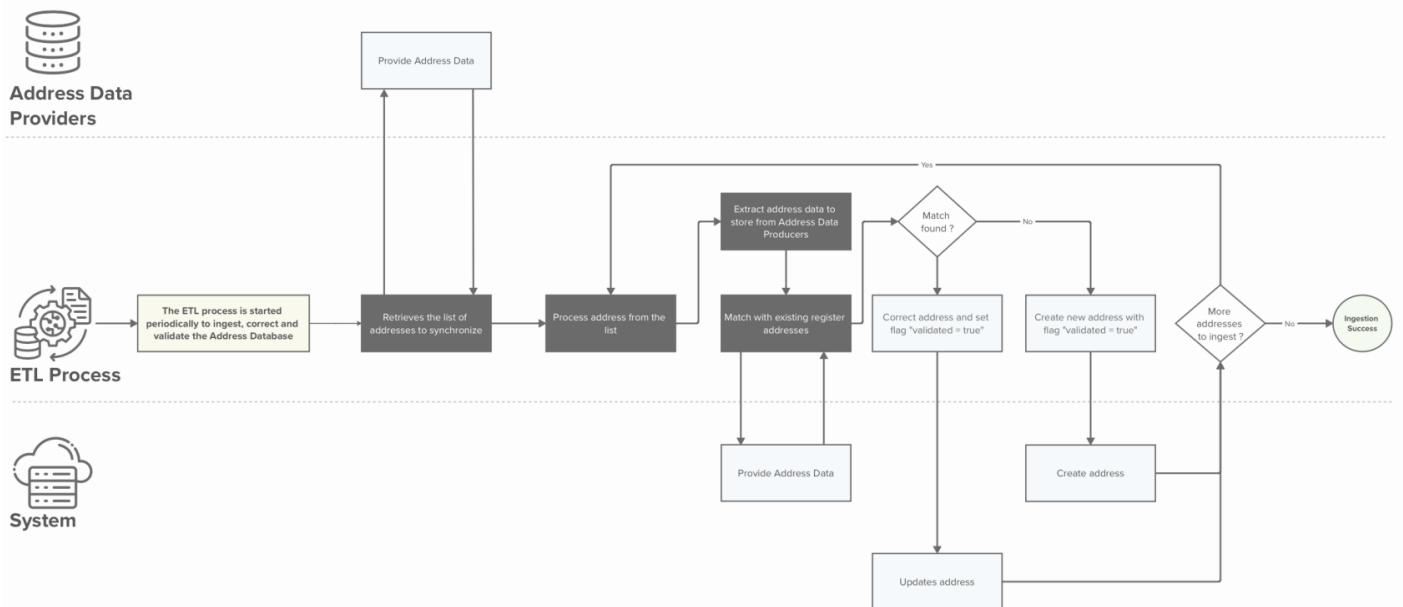
The ETL process that is responsible for ingesting data from one or multiple trusted external Address Data Providers, is a key element of the RNCV. The ETL is responsible for:

- Ingest new addresses present on the Address Data Providers.
- Correct existing addresses
- Validate addresses ingested by the Data producers

You will find below a high level representation of the ingestion process. The exact algorithms used by the ETL process are out of the scope of this project and therefore are abstracted as a black box in the process description

ETL Process

ETL Address Ingestion Process



Name	Address ingestion via the ETL process
Purpose	Ingest and validate address data to maintain the quality of the RNCV's address database database to the highest standards
Linked user stories	4.67. ETL - Retrieve addresses 4.68. ETL - Create and update addresses
APIs used	GET /etl/addresses POST /etl/addresses GET /etl/addresses/<address-id> PUT /etl/addresses/<address-id> PATCH /etl/addresses/<address-id>
Scope	This process handles the ingestion of addresses into the RNCV's address database. It also handles the correction and validation of existing address data. This exact algorithm used by the ETL process is out of scope of this process.
Roles	ETL, System
Input	- Addresses from the Address Data Providers - Algorithm for the address data consolidation

Name	Address ingestion via the ETL process
Output	- Consolidated and up to date RNCV address database

Detailed Process description

Main process

Step	Description	Actor(s)	Input(s)	Output(s)	Decision points
1	The ETL process is periodically triggered	ETL	-	-	
2	The ETL process retrieves the data to synchronise from the Address Data Providers	ETL	-	- addresses to synchronise	
3	The ETL process process one address from the addresses to synchronise (could be done in parallel)	ETL	- addresses to synchronise	- next address to synchronise	
4	The ETL process extracts the address information to be stored in the RNCV address database	ETL	- address to synchronise	- address information to be ingested	
5	The ETL process searches for address matches in the RNCV address database	ETL	- address information to be ingested	- address present in the RNCV address database if any	
6	The ETL process checks if the address is present in the RNCV address database	ETL	- address present in the RNCV address database if any	- yes / no	If the address is present: Go to step 7 Else: Go to secondary process S.1.
7	Correct address information and set the flag "validated = true"	ETL	- RNCV address	- Corrected RNCV address with flag "validated = true"	
8	System applies the address update	System	- Corrected RNCV address with flag "validated = true"	- Corrected RNCV address with flag "validated = true"	

Step	Description	Actor(s)	Input(s)	Output(s)	Decision points
9	The ETL process checks if more addresses need to be synchronised	ETL	- addresses that still need to be synchronised	- yes / no	If there are still addresses to be synchronised: Go to step 3 Else: Go to step 10
10	The ETL process terminates successfully	ETL	-	-	

Secondary Processes

S.1. Address does not exist in the RNCV address database

Step	Description	Actor(s)	Input(s)	Output(s)	Decision points
1	The ETL process creates a new Address with the flag "validated = true"	ETL	- address information to be ingested	- RNCV address to be created	
2	The system creates the given address	System	- RNCV address to be created	- RNCV address created	Go to Main process step 9

Additional Information

Error processing during the ETL process

If an error occurs during the ETL process (internal error, or error while using an external API), the system should log the error and process the next address. An error triggered during the processing of one address should never interrupting the ETL process for subsequent addresses, except if it is a system wide error, that would prevent all addresses from being processed.

Revision #2

Created 20 October 2025 08:55:43 by Sergio Sousa

Updated 25 November 2025 10:17:09 by Sergio Sousa