



**CARTOGRAPH AI**

WHITEPAPER  
CARTOGRAPH SOLUTIONS

# Contents

- I. Introduction.....3
  - A. Problem .....3
  - B. Solution .....3
    - 1. Profiling and Discovery .....3
    - 2. Mapping and Transformation .....3
  - C. Use Cases.....3
    - 1. New Implementations .....4
    - 2. Additional Data Sources .....4
    - 3. Data Assessments.....4
    - 4. Reverse Engineering .....4
    - 5. Governance and Lineage .....4
- II. Functionality.....5
  - A. Modules.....5
    - 1. Discover.....5
    - 2. Map .....6
    - 3. Transform .....7
    - 4. Governance.....7
  - B. User Interface .....7
    - 1. Data Source Configuration.....7
    - 2. Target Application Configuration .....8
    - 3. Discover Explorer .....9
    - 4. Map Matrix .....10
    - 5. Transform Designer .....10
    - 6. Governance Administration .....11
- III. Conclusion .....12

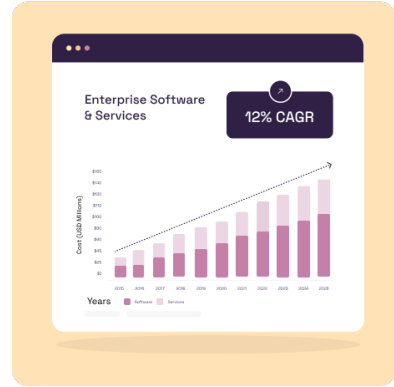
# Version Control

Version	Author	Details	Date
1.0	Patrick Robak	Version 1	12/5/2024
1.1	Patrick Robak	Addition of graphics	7/16/2025
1.2	Patrick Robak	General revisions	8/13/2025
1.3	Patrick Robak	General revisions	2/10/2026
1.4	Patrick Robak	Revised Governance sections	2/24/2026

# I. Introduction

## A. Problem

LexisNexis Risk Solutions released its annual report detailing the “True Cost of Financial Crime Compliance”. The 2023 study revealed that FCC costs companies \$206.1 billion globally, and the driving force behind the surge is the rising price of expertise and technology. This problem is not unique to Anti-Money Laundering, or even Financial Institutions – it applies to all Enterprise Software and Services across all domains and businesses.



While combating rising costs, technology teams are also contending with common organizational challenges:

- People – Resource constraints and unclear data ownership
- Processes – Manual data wrangling and nascent data governance
- Technology – Use of redundant tools and limited architecture discipline
- Data – Lack of trust in siloed data and limited awareness of data assets

## B. Solution

CartographAI is an automated data integration and management engine that supercharges subject matter expertise with artificial intelligence. Core functionalities include:

### 1. Profiling and Discovery

- a. CartographAI analyzes, reviews, and summarizes source data
- b. Understands structure, content, and interrelationships
- c. Assigns meaning to tables (ex. Transactions, Accounts, Customers, etc.)
- d. Assigns meaning to attributes (ex. Transaction Amount, Account Number, Customer Type, etc.)
- e. Flags quality, accuracy, completeness, consistency, timeliness, and accessibility issues

### 2. Mapping and Transformation

- a. CartographAI extracts relevant data from the source
- b. Transformations adhere to target data types, constraints, and dependencies
- c. Data is loaded according to the specified timing and frequency
- d. Detailed documentation is automatically generated for transparency and traceability
- e. Mappings requiring manual intervention are flagged for review
- f. Governance framework ensures data quality does not degrade over time

## C. Use Cases

Compelling use cases for CartographAI continue to emerge as the product develops and is socialized with senior industry insiders facing unique challenges.

## 1. New Implementations

The implementation of a new software application represents perhaps the most natural use case for CartographAI. Unknown and poorly documented sources must be mapped to an entirely new target data model within the constraints of an aggressive project timeline. CartographAI accelerates this effort with a high degree of confidence, allowing team members to focus on other complex deliverables.

## 2. Additional Data Sources

A smaller effort but more frequent occurrence, the addition of new data sources is largely automated using CartographAI. Bring a new business line or product online faster while ensuring compliance with data security requirements.

## 3. Data Assessments

With data being the foundation upon which models are built, an increasing number of stakeholders are taking a microscope to the end-to-end data quality, completeness, governance, and lineage feeding those systems. CartographAI breaks down knowledge silos and provides unprecedented transparency to model owners, validators, auditors, and regulators.

## 4. Reverse Engineering

The data mappings, transformations, and lineage feeding our business applications are often poorly documented and outdated. Legacy artifacts in the form of Word documents and Excel spreadsheets are stored in SharePoint and Network Drives where they become increasingly obsolete over time. CartographAI can reverse engineer existing code and mapping specifications and transform them into living, inspectable assets. Documentation is always current and defensible.

## 5. Governance and Lineage

Having end-to-end data mappings and transformations within CartographAI provides the foundation for the Data Governance and Lineage features. Rule-based data quality and privacy rules can be applied to input (source) and output (target) data to ensure issues are swiftly identified and remediated. The complete lineage of any source or target entity can be easily traced throughout the entire data journey, from the highest entity level down to a specific attribute.

## II. Functionality

### A. Modules

CartographAI is comprised of four core modules: Discover, Map, Transform, and Governance.

#### 1. Discover

The Discover module of CartographAI rapidly accelerates the discovery, analysis, and profiling of unknown and poorly documented data sources.

It connects to and interprets data sources of many different types, including structured and unstructured data. These include, but are not limited to RDMS, NoSQL, Data Lakes, Lakehouses, CSV, XML, JSON, Parquet, and other file types. CartographAI does not store any source data, only the relevant metadata and the insights/outputs it produces.



After connecting to a source, Discover consumes various inputs to gain deeper insights into the nature of the data. The more inputs available to the Discover module, the more accurate the outputs are expected to be. However, input data quality is also an important consideration, as poor input quality could adversely affect results. Discover inputs include:

- a. Source table names, schemas, and indexes
- b. Source attribute names, data types, lengths, defaults, and constraints
- c. Source attribute values
- d. Source metadata
- e. Source documentation

After evaluating the available inputs, the Discover module outputs several key artifacts that detail the data source. Discover outputs include:

- a. High-level functional descriptions of each table and each attribute
- b. A detailed data dictionary
- c. An Entity Relationship Diagram (ERD)
- d. A data profiling report

In addition to the aforementioned outputs, the Discover module also offers the following features:

- a. Confidence score: All insights generated by the Discover module are measured against a configurable confidence score. Any output that does not meet the predetermined threshold is flagged as such and either requires modified module inputs or manual verification.
- b. Co-pilot: The Discover Co-pilot is an AI-powered virtual assistant. Users can interact with the co-pilot to clarify Discover insights and drill deeper into source databases, tables, attributes, and interrelationships.
- c. Governance exceptions: The Discover module has the ability to flag data governance exceptions, either in the form of data quality or data privacy concerns. This helps ensure

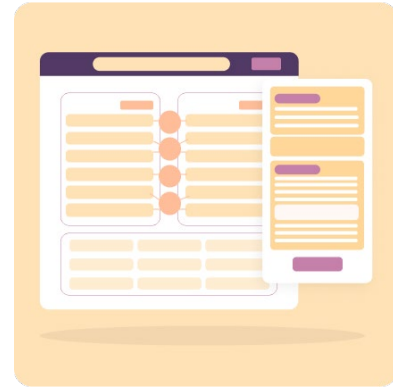
any issues or changes in the source data are detected early, before being transformed and passed to the target application.

## 2. Map

The Map module of CartographAI leverages the Discover module output along with a knowledge base of mapping rules to propose detailed source-to-target mappings for any target FCC application.

CartographAI will include mapping rules for all significant Financial Crime Compliance applications, beginning with the most prevalent. The knowledge base will be expanded and improved over time.

After evaluating the available inputs, the Map module outputs detailed and exportable source-to-target mapping specifications for each critical target attribute.



In addition to the aforementioned outputs, the Map module also offers the following features:

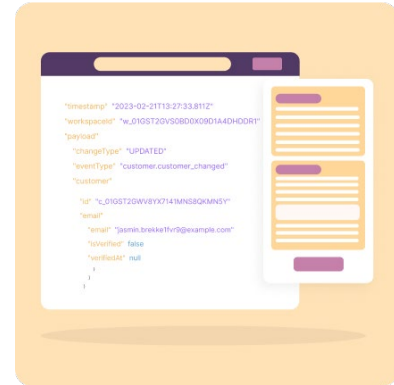
- a. **Manual review/override:** Each mapping produced by CartographAI can be manually reviewed in the Map Matrix UI. The mapping can either be approved or overwritten, if deemed necessary. All such actions would be fully audited.
- b. **Sample output:** Each mapping is accompanied by a real-time output sample to visualize the effects of any applied transformations. This assists the user in validating the mappings and any necessary changes.
- c. **Output validation:** Automated scripts that validate the mapping output to ensure it conforms to the necessary data types, constraints, and dependencies.
- d. **4-eye review:** A validation that requires at least 2 users to approve a mapping. These validations can be applied to specific target attributes, or across all mappings.
- e. **Confidence score:** All mappings generated by the Map module are measured against a configurable confidence score. Any output that does not meet the predetermined threshold is flagged as such and either requires modified module inputs or manual verification.
- f. **Reverse engineering:** Rather than use Discover insights and the rule library as inputs to produce source-to-target mappings, use existing ETL code and scripts as inputs to reverse engineer existing mappings. This is useful in cases where the current transformations and pipelines are poorly documented but need to be validated. The reverse mappings can also be compared to those CartographAI would typically generate to identify potential gaps for remediation.

### 3. Transform

After generating detailed data mapping specifications, Transform consumes those instructions as an input to autonomously generate ETL code/pipelines that can be executed manually or through a job scheduling tool.

ETL pipelines come in different forms, all of which can be accommodated by CartographAI. Including, but not limited to SQL, Python, Apache Airflow, Apache Spark, Google Composer and more.

In order to productionize ETL pipelines, CartographAI seamlessly integrates with data orchestration systems in the client's environment.



### 4. Governance

The CartographAI data governance framework is a set of rules and processes that manage data within the application. These rules can be applied at the input or output levels and are accessible from the Modules drop-down menu within a project, along with Explorer, Matrix, and Designer.

Data Governance rules fall into one of two categories: Quality or Privacy. Data Quality rules test for potential issues with data uniqueness, consistency, completeness, referential integrity, etc. Data Privacy rules flag tables and attributes that may contain Personally Identifiable Information (PII), Payment Card Industry (PCI), or Employee (EMP) data.

#### B. User Interface

CartographAI features a specially designed user interface to configure and interact with each core module. The main components of the UI include:

##### 1. Data Source Configuration

The data source interface is a way to connect CartographAI to data sources of various types, typically by entering credentialed information like username, password, hostname, port, etc. The interface saves connection details for future use.

×

Source Type

Database    File

Source Name

Database Type

Database Name

Host

Port

Username

Password

## 2. Target Application Configuration

The target application interface is a way to filter the mapping rule library to only map those fields within the scope of each specific project. The user specifies the target application, modules, rules, and even field usage.

×

Basic DetailsCollaboratorsConfigure SourceConfigure Mapping

Application

NICE Actimize ▼

Modules

AML-SAM (Suspicious Activity Monitoring) ▼

Rules

Select the rules to be implemented ▼

Usage

Select rules first ▼

Create

### 3. Discover Explorer

The Explorer is the primary screen of the Discover module, displaying all the relevant outputs in a user-friendly and interactive manner.

**Explorer accelerates source data analysis**

**LLM generates Entity Relationship Diagram**

**High-level source table and field descriptions**

**Dictionary identifies object type, keys, and constraints**

**Detailed data profiles highlight data quality**

**Top 10 records from each source table**

TRANSACTION ID	ACCOUNT ID	CONTRA ACCOUNT TYPE	CONTRA ACCOUNT ID	CONTRA BANK ID	CONTRA BANK NAME	CURRENCY CODE	DIRECTION	SOU
2023813-21	253515602	EXT	135413929	CMCIFRPP	CMCI	USD	C	
2023615-22	416358625	EXT	122593998	ZUNOCZPP	ZUNO	USD	C	
20231130-23	666212884	EXT	416358625	DXIADEBB	DXIA	AUD	C	
2023217-24	200917270	EXT	477237000	GEBABEBB	GEBABEBB	AUD	C	
20231028-25	296903536	EXT	639668286	ABNACHZ6	ABNA	CNY	D	

**TRANSACTION**

Description: This table records all transactions occurring within the system. It includes details such as transaction ID, account information, and amounts. Each transaction is linked to various types and sources for comprehensive tracking.

Dictionary: Dictionary

Profiles: Profiles

Columns: 12, Rows: 1000

Primary Keys: 1, Foreign Keys: 5

Column Overview:

Name	Unique %	Missing %
TRANSACTION_ID	100%	0%
ACCOUNT_ID	10%	0%
CONTRA_ACCOUNT...	0.2%	0%
CONTRA_ACCOUNT...	10%	0%

## 4. Map Matrix

The Matrix is the primary screen of the Map module, displaying all the relevant mappings in a user-friendly and interactive manner.

**Matrix automates mapping specifications**

**LLM recommends source to target mappings**

**High-level target table and field descriptions**

**Simulation shows users output in real-time**

**Editable pseudo-logic for recommended mappings**

**Overall and table-level mapping progress**

**Show users how each table and field is used in the application**

**Full audit history of every system and user action**

**Source: ACCOUNT**

ACCOUNT\_NUMBER  
NUMBER\_OF\_ACCOUNT\_TITLE\_LINES  
ACCOUNT\_REGISTRATION\_LINE\_1  
ACCOUNT\_REGISTRATION\_LINE\_2  
ACCOUNT\_REGISTRATION\_LINE\_3  
ACCOUNT\_REGISTRATION\_LINE\_4

**Target: ACCOUNTACCOUNT\_KEY**

ACCOUNT\_CATEGORY\_CD  
ACCOUNT\_CLA  
ACCOUNT\_...  
ACCOUNT\_KEY  
ACCOUNT...  
ACCOUNT...  
ACCOUNT...

account_type_cd	account_number	account_first_name	account_last_name	account_middle_name	account_name	account_c
3	178060873	Morris	Letler	Jr.	Morris Letler Jr.	MORRIS
R	796159943	Arlie	Cummerata	Null	Arlie Cummerata	ARLIE
R	135413929	Katelynn	Beier	DDS	Katelynn Beier DDS	KATELYNN
T	875625879	Joe	Berge	Null	Joe Berge	JOE
T	477555925	Wilbert	Wintheiser	Null	Wilbert Wintheiser	WILBERT

**ACCOUNTACCOUNT\_KEY**

Description: Unique key across all tenants for the account. An institution could have accounts from different sources with the same account number which will need to be made unique across all sources such that it does not duplicate any other account\_key in any other tenant.

Mapping: ACCOUNTACCOUNT\_KEY

Progress: Progress

Usage: Usage

History: History

Simulate

## 5. Transform Designer

The Designer is the primary screen of the Transform module. It is an easy-to-use web interface where users can create, modify, and manage processes.

Designer turns pseudo-logic into exportable code

LLM converts pseudo-logic text into code

High-level target table and field descriptions

Editable pseudo-logic for recommended mappings

Full audit history of every system and user action

Generate code in programming language of choice

```

1 SELECT
2   CONCAT(src.source_system, '-', src.account_id) AS ACCOUNT_KEY,
3   CONCAT(src.source_system, '-', src.contra_account_id) AS OPP_ACCOUNT_KEY,
4   CONCAT(src.source_system, '-', src.transaction_type) AS TRANSACTION_CODE_CD,
5   CONCAT(src.source_system, '-', src.transaction_id) AS TRANSACTION_KEY,
6   src.amount AS ACCT_CURR_AMOUNT,
7   src.currency_code AS ORIG_CURR_AMOUNT,
8   src.source_system AS SOURCE_SYSTEM_CD,
9   TO_CHAR(NOW()::timestamp, 'YYYY/MM/DD') || ':00:00:00' AS BATCH_DATE_TIME,
10  TO_CHAR(
11    src.transaction_date_time::timestamp,
12    'YYYY/MM/DD'
13  ) AS EXECUTION_LOCAL_DATE_TIME,
14  CASE
15    WHEN src.contra_account_type IS NOT NULL THEN COUNT(src.contra_account_id) + 1
16    ELSE 0
17  END AS NUM_INTERMEDIARIES,
18  CASE
19    WHEN src.contra_account_type = 'EXT' THEN REGEXP_REPLACE(src.contra_account_id, '[^A-Za-z0-9]', '')
20    ELSE NULL
21  END AS OPP_ACCOUNT_NUMBER,
22  CASE
23    WHEN src.contra_account_type = 'EXT' THEN src.contra_bank_id
24    ELSE NULL
25  END AS OPP_ORGANIZATION_KEY,
26  CASE
27    WHEN src.account_id IS NOT NULL THEN 'Originator Name'
28    ELSE NULL

```

TRANSACTIONS

Description

This entity is used to capture all monetary transactions associated with the accounts that do not meet the criteria listed to be populated on trades or transfers tables. These include transactions like cash deposits and withdrawals, monetary instruments deposits and withdrawals, credit card payments, etc.

Mapping

History

Generate

## 6. Governance Administration

The CartographAI data governance framework is a set of rules and processes that manage data within the application. The purpose of the framework is to ensure the integrity, security, and compliance of data, and to establish a standard for how data is collected, organized, stored, and used. This makes it easier to streamline and scale data governance, maintain policy and regulatory compliance, democratize data, support collaboration, and build trust.

Classify rules as Data Quality or Data Privacy

Specification of rule type

Configuration of Data Governance rules

High-level rule descriptions

Activation and deactivation of rules

Rule ID	Class	Type	Description	Status
DG1	Quality	Duplicate	Fields whose values are highly duplicative of another field in the same or different table (excluding primary and foreign keys).	Active
DG2	Quality	Inconsistent	Fields whose values are inconsistent with the description of the field.	Inactive
DG3	Quality	Outlier	Fields with outlier values that differ greatly from the other values.	Active
DG4	Quality	Incomplete	Fields with missing or partially missing values.	Active
DG5	Quality	Outdated	Date fields where the max value is significantly older than the current system date.	Active
DG6	Quality	Distinct	Categorical fields with a newly added distinct value.	Inactive
DG7	Privacy	PII	Field may contain Personally Identifiable Information data.	Active
DG8	Privacy	PCI	Field may contain Payment Card Industry data.	Active
DG9	Privacy	Employee	Field may contain Employee data.	Active

### III. Conclusion

Data management is a crucial component of an effective FCC program. CartographAI reduces implementation and maintenance costs without sacrificing quality.

The solution accelerates data mapping, offering financial institutions a marked improvement on the time-to-value of their investments and instills confidence in the underlying data that is often severely lacking.

The solution also transforms legacy and artifacts into living, inspectable assets. This means documentation is always current and defensible, eliminating the scramble before an audit or regulatory exam.

Reimagine your data journey now.