FDIC Semantic Augmentation Project
2025 OMG Semantic Augmentation Challenge
Author: Kemar Lee

Project Summary

This submission tackles the semantic enrichment of FDIC institutional bank data. I focused on turning a flat CSV into something machines can understand and reuse using RDF metadata, standardized URIs, and a simple conceptual model. The goal was to make the data more useful without overcomplicating the approach.

Objectives

- Link each column in the dataset to a real-world semantic identifier
- Build a metadata file (TTL) that clearly explains what each field means
- Create a simple but standards-aligned model of the data using UAF concepts
- Deliver something lightweight, valid, and usable

Approach

I manually mapped each column in the FDIC dataset to a term from one of three sources:

- **FIBO** — for financial fields
- **Schema.org** — for general descriptors like name, city, etc.
- A small custom **fdic:** namespace — for FDIC-specific identifiers and categories

Each mapping includes a datatype, label, and short description. These are stored in both a mapping_table.csv (for human review) and a metadata.ttl file in RDF/Turtle format.

Units like “USD” and “%” are noted using a simple ex:unit property. I avoided adding complexity that would make the metadata harder to interpret or reuse.

Conceptual Model (Cameo EA 2021x)

I built the conceptual model using the limited trial version of Cameo Enterprise Architect (2021x). Since the trial restricts full access to UAF stereotypes, I worked within what's available:

- Used a Standard block called **Bank** as the top-level concept

- Created 15 connected elements (matching the dataset columns) using **Directed Aggregation**
- Aligned the model visually and structurally with the metadata

It's not a full UAF Taxonomy diagram, but it follows the same logic and grouping principles.

Data Transformation

To validate the mappings, I applied them to a sample dataset:

- Replaced the original headers (e.g., Cert, City, Net Income) with semantic URIs
- Saved the transformed file as transformed_data.csv
- Data values remained unchanged — only the headers were updated for semantic clarity

A lightweight internal script was used, but it's not included since the process is already visible in the submitted files.

Roadblocks & Workarounds

Tooling: Cameo EA trial was limited — no class diagrams, no full UAF profile. I improvised with the Standards Taxonomy view and kept the structure clean and consistent.

Data Format: Some fields (e.g., percentages, dates) needed normalization. I used standard XSD types to keep things consistent.

URI Alignment: Not every field had a perfect match in existing ontologies. I made careful calls where needed, aiming for clarity and reuse.

Final Notes

I designed this to be standards-compliant without over-engineering. It's focused, understandable, and reflects what the challenge actually asked for: meaningful metadata, useful structure, and alignment with real-world ontologies and modeling frameworks.
