**FDIC Semantic Transformer App – Business Documentation**

The **FDIC Semantic Transformer** is a Streamlit-based web application that automates the semantic enrichment of the FDIC Summary of Deposits (SOD) CSV dataset. It allows non-technical users to upload a raw CSV file, instantly transform it into machine-readable RDF format, and validate it against regulatory rules using SHACL constraints.

**🧩 Key Features and Functional Benefits**

**1. 📁 CSV Upload with Intuitive Interface**

* **What it does:** Users upload the FDIC-insured banks dataset in CSV format.
* **Business value:** No coding required—this allows data stewards, analysts, and domain experts to begin semantic processing without needing a developer.



**2. 🔄 Automatic RDF Transformation**

* **How it works:**
On clicking "Transform to RDF," the app:
	+ Loads the ontology (fdic\_ontology\_fixed.ttl)
	+ Maps CSV columns using skos:altLabel
	+ Converts fields to RDF triples using proper data types and URIs
	+ Includes metadata using **DCAT** and **VoID** vocabularies (e.g., issue date, record count)
* **Business value:**
	+ Transforms ambiguous spreadsheet data into **linked data** suitable for integration, AI, and reporting.
	+ Preserves schema consistency across dataset versions.
	+ Automatically handles tricky date fields like ACQDATE, ESTYMD, RUNDATE—including Excel serial dates and inconsistent formats.

**3. 📊 Preview Original and Transformed Data**

* **UI Features:**
	+ Displays the first record from the CSV
	+ Shows RDF triples (in Turtle syntax) generated for that record
* **Business value:**
	+ Ensures **transparency and traceability** for validation.
	+ Helps users understand how their raw data is being interpreted semantically.





**4. ✅ Built-in SHACL Validation**

* **How it works:**
Users select how many records to validate (1, 2, 5, 10, or all), then click "Validate with SHACL."
	+ The app:
		- Extracts a clean RDF subset
		- Loads SHACL rules from fdic\_shacl\_constraints\_optimized.ttl
		- Runs pySHACL to check conformance
		- Displays validation results, including specific issues per record
* **Business value:**
	+ Guarantees **regulatory compliance** by enforcing:
		- Correct state codes and names
		- Proper data types (e.g., date formatting, integer flags)
		- Required fields (e.g., type, main office indicator)
	+ Makes the data **fit for automated pipelines and reporting systems**.



**5. ⬇️ Download Options for RDF and SHACL Reports**

* Users can:
	+ Download the full RDF graph in Turtle format
	+ Download the SHACL validation report
* **Business value:**
	+ Facilitates **audit trails**, **interoperability**, and **data exchange** with external systems or regulators.



**🧠 Behind the Scenes: What the App Understands**

The app smartly connects CSV column names to ontology predicates by:

* Loading all skos:altLabel annotations from the ontology
* Dynamically resolving each field's expected **data type** (e.g., xsd:date, SKOS concept)
* Handling complex input formats like Excel serial dates or inconsistent date delimiters

This makes it **future-ready** for changes in CSV format or content.

**💼 Summary of Business Impact**

|  |  |
| --- | --- |
| **Capability** | **Business Value** |
| No-code UI | Empowers non-technical teams to perform semantic modeling |
| Semantic RDF generation | Enables integration, querying, and explainability |
| Regulatory rule validation | Ensures data quality and compliance |
| Metadata tagging with DCAT/VOID | Supports data cataloging, sharing, and reuse |
| Repeatability across versions | Guarantees schema consistency for quarterly FDIC reports |
| Extensible to JSON/XML inputs | Future-proof for multi-format support |