OFFICE OF FINANCIAL RESEARCH
STAFF DISCUSSION PAPER

18-01 | June 27, 2018

An Ontology of Ownership and Control Relations of Bank Holding Companies

Liju Fan
Office of Financial Research
liju.fan@ofr.treasury.gov

Mark D. Flood
Office of Financial Research
mark.flood@ofr.treasury.gov

An Ontology of Ownership and Control Relations for Bank Holding Companies[1]

Liju Fan
Office of Financial Research
liju.fan@ofr.treasury.gov

Mark D. Flood
Office of Financial Research
mark.flood@ofr.treasury.gov

Abstract

We consider the challenges and benefits of ontologies for information management for regulatory reporting from bank holding companies (BHCs). Many BHCs, especially the largest and most complex firms, have multiple federal supervisors who oversee a diverse array of subsidiaries. This creates a federated data management problem that disperses information across many firms and regulators. We prototype an ontology for the Federal Reserve's public National Information Center (NIC) database. The NIC identifies all BHCs, their subsidiaries, and the ownership and control relationships among them. It is a basic official source on the structure of the industry. A formal ontology can capture this expert-curated knowledge in a coherent, structured format. This could assure data integrity and enable non-experts to more readily integrate and analyze data about complex organizations. We test the design and development of federated prototype ontologies in the Web Ontology Language (OWL) to provide and integrate the NIC data with precise semantics for transparency and consistency. Our preliminary results indicate that this is feasible in practice for data search and analysis, and that the ontologies can facilitate semantic integration and improve the integrity of data and metadata.

Keywords

Financial regulation; data integration; knowledge representation; ontologies; Web Ontology Language (OWL); data integrity; bank holding companies.

[1] Views expressed in this paper are those of the authors and do not represent the official views or position of the OFR or the U.S. Department of the Treasury. A version of this paper was presented at the SIGMOD/DSMM 2018 conference in Houston. The authors thank Alicia Marshall for expert legal advice, Somgnadeth Seumsouk for expert dataset processing, and Louiqa Raschid for her encouragement.

1. Institutional Context

In the United States, bank holding companies (BHCs) are a formal legal entity type that can provide tax and regulatory advantages compared with a free-standing bank. Financial firms commonly organize as BHCs, with a "high-holder" BHC at the root of the ownership tree. The high-holder then owns — directly or indirectly – one or more banks and perhaps non-banks as subsidiaries. The Federal Reserve Board (Fed) regulates BHCs in the United States. The Fed maintains the National Information Center (NIC) database to track BHCs and their subsidiaries. The Office of Financial Research (OFR) uses the NIC in support of its mandates to standardize regulatory data collections and research financial stability issues.

In this paper, we outline an approach to capturing the key semantic information in the NIC about BHCs, subsidiaries, and their ownership and control relationships. The NIC database contains over 100 fields, and the data dictionary only sketches their definitions. Comprehensive knowledge lies scattered across hundreds of laws, regulations, and official pronouncements with interpretations often accessible only to experts. We capture this detailed knowledge in an ontology that non-experts can use to analyze BHCs more effectively. This extends earlier efforts [8] to apply structured semantics to financial regulatory filings, in which we introduced a regulatory form ontology and an upper ontology and relations (UOR). We redeploy and extend the UOR here.

1.1 Structure of Bank Holding Companies

BHCs come in many shapes and sizes. The simplest possible structure is a single bank owned entirely by one BHC, known as a "unitary BHC." In contrast, the largest financial firms, such as the global systemically important banks (G-SIBs), may have hundreds or even thousands of subsidiaries below the high-holder [12].

A BHC is only one of several types of holding company. Other examples include *Savings and Loan Holding Company* (SLHC) and *Financial Holding Company* (FHC). The differences between BHCs, SLHCs, and FHCs are formal legal distinctions that evolve through time with new laws and regulations issued. A more specialized case is the *Intermediate Holding Company* (IHC), which is established by a "foreign banking organization (FBO)" to hold its ownership interest in U.S. subsidiaries, per Subpart O of the Fed's Regulation YY. The economics literature frequently uses "BHC" as a shorthand to refer to all holding company types.

Understanding these elaborate structures is an important practical challenge. For example, when a BHC wants to acquire a new bank subsidiary, or an existing bank wants to establish new branches, regulators analyze the competitive implications to ensure the action will not create excessive concentration in a particular geographic area [6]. This requires knowing which high-holders control each of the subsidiaries in the relevant jurisdiction. When an institution fails, regulators must resolve the firm quickly, selling off or closing certain subsidiaries, or putting them into conservatorship [5]. For very large BHC hierarchies, the intricacy of ownership can itself contribute to a firm's opacity, hindering the ability of investors and regulators to understand the risks [13].

In the United States, oversight of individual BHC subsidiaries is assigned to one or more of the regulators [14]. For entities with multiple regulators, a primary regulator is designated to coordinate among them.

Under its Regulation Y, the Fed collects quarterly financial data from BHCs (reported on Form FR Y-9) and regulates their activities, even in cases where another agency is the official primary regulator.

1.2 Ownership and Control Concepts

Subsidiaries' relationships within a BHC can take many forms. For example, subsidiaries typically issue ownership shares, allowing one subsidiary to have multiple owners, each with a partial stake. We call the owning/controlling entity the "parent" and the owned/controlled entity the "offspring." Direct ownership relationships can typically be composed transitively to generate indirect relationships; for example, if ***A*** owns ***B*** directly, and ***B*** owns ***C*** directly, then ***A*** owns ***C*** indirectly [26].

Analysts often distinguish ownership from control relationships [1]. Shareholders' voting rights define control over a firm — for example, rights to pick the firm's board of directors. Ownership, on the other hand, makes shareholders residual claimants on the firm's resources, including cash flow rights to any profits that are not otherwise obligated to creditors. When a single owner holds 100 percent of the shares, ownership and control become indistinguishable; this is typical for BHC subsidiaries. However, there are many ways to create a wedge between ownership and control [7] [22]. For example, at a firm issuing only one class of stock with a one-share/one-vote rule, a shareholder with a 51-percent stake would have complete effective control (simple majority voting rights) but just over half of the cash flow rights.

Other nuances include issuance of a separate class of nonvoting preferred stock; or convertible bonds (i.e., debt) that bondholders can exchange for a predetermined amount of equity shares; or managerial stock options — which lack voting and cash flow rights — but which can be exercised to purchase full-fledged equity shares. Beyond formal distinctions, there are practical problems, such as shareholders who neglect to vote because their stake is too small to affect the outcome, or the agency problems associated with shareholders' difficulties in monitoring or controlling the firm's managers, who run the business from day to day[1][18].

The NIC captures ownership and control in the ***Relationships*** table. For example, the NIC records the exact ownership percentage in the `PCT_EQUITY` field if the subsidiary is a banking company (`EQUITY_IND=1`), but an approximate percentage when the subsidiary is a nonbanking company (`EQUITY_IND=2`). The NIC treats bank branches as components of the bank, which do not participate directly in ownership and control relationships with other subsidiaries of a holding company. Nor do branches have a separate "regulatory or reporting relationship" with the Fed. The NIC tracks branches in a separate table, linking each branch to its headquarters office by a foreign key to the entity that does participate in ownership and control relationships.

2. Methodology

2.1 National Information Center Database

To support supervision, the Fed launched the NIC in 1988 as an information resource to be shared across the Federal Reserve System. The NIC has evolved into its current form as a "central repository of information about all U.S. banking organizations and their domestic and foreign affiliates" [10]. The Fed updates the NIC daily for internal use, and publishes quarterly snapshots of most of the NIC, including ownership and control relationships, through the Federal Financial Institutions Examination Council (FFIEC), which coordinates among the regulators with a focus on depository institutions. We work with the FFIEC's public version of the NIC, which is downloadable as five separate files [9]:

- ***Attributes-Active*** – details on going-concern entities,
- ***Attributes-Closed*** – details on closed entities,
- ***Relationships*** – details of ownership and control relations among BHC subsidiaries, and
- ***Attributes-Branches*** – details for branches that are subunits of an (active or closed) reporting entity, and
- ***Transitions*** – key lifecycle events, such as mergers.

We focus on the first and third items, covering the NIC's structure elements describing subsidiaries of every BHC under the Fed's jurisdiction, and the ownership and control relationships that connect them. The Fed assigns an RSSD ID to identify each entity in the NIC. RSSD is an abbreviation for "Research, Statistics, Supervision and Regulation, and Discount and Credit Database" [14, p. 54].) RSSD IDs are stable over time, typically tracking a firm from its formation until it is extinguished by acquisition or failure. In some cases, a BHC converts a subsidiary (a distinct legal entity) into a branch of another subsidiary; in many cases, the branch (with its RSSD ID) will remain in the NIC, even though it is no longer a free-standing legal entity.

Because the Fed's legal jurisdiction is limited to the United States, the NIC tracks only firms that have U.S. operations, including firms with headquarters abroad. The primary source of raw information to populate the NIC structure elements comes on the Fed's reporting Form FR Y-10 [4]. Federal Reserve Banks and the Board of Governors curate the NIC database using the RSSD system, a shared application [2]. The Fed publishes a data dictionary describing the NIC [3]. An important field for our analysis is the controlled vocabulary of entity types; each entity in the database is labeled with exactly one `ENTITY_TYPE` at any point in time.

The NIC tracks the status of entities over time, recording every merger, failure, or change in location or charter type. In many cases, the official legal name for a particular entity can change while the RSSD ID remains unchanged. For example, consider the evolution of the entity with RSSD ID #1120754 [13, Table 1]. This entity is now known as *Wells Fargo & Company*, but it began its existence in 1929 as *Northwest Bancorporation*. Over the years, this entity has kept its RSSD ID but has changed its official name twice (1983 and 1998), changed its headquarters location twice (1984 and 1998), and changed its entity type twice (1956 and 2000). The Fed ultimately decides whether a particular organizational change is substantial enough to justify the issuance of a new RSSD ID versus simply changing the database attributes associated with the existing RSSD ID.

The choice of which entities to reify and identify in the database is an important ontological commitment. For the NIC, the guiding principle is to record in one place key facts about all "depository institutions, regulated and select non-regulated institutions, and other institutions that have a regulatory or reporting relationship" with the Fed [2, p. 8]. Other databases of financial entities have other priorities and make other data modeling commitments. For example, the Legal Entity Identifier (LEI) repository focuses on identifying contractual counterparties — the individual subsidiaries, in contrast to their high-holder parent — that participate as obligors (and obligees) in financial transactions [15]. Although the LEI is codified in an international standard [16] and managed through a robust governance framework, it faces challenging semantic nuances in practice. For example, most natural persons have legal standing to act as obligors, but they are typically ineligible for an LEI unless they are acting in a "business capacity" [19]. An individual bank depositor ordinarily could not get an LEI, but the same person might be eligible as the sole proprietor of a business.

2.2 Practical Applications of an Ontology

Two characteristics of the NIC data suggest an ontology will be beneficial for the quality of both the data and metadata. First, the domain of financial firms is in constant flux; an ontology's rigorous structure can help users maintain the internal coherence of NIC metadata across updates. Current processes emphasize instead expert (human) curation to keep up with ongoing changes. Second, the NIC is a critical input to financial regulation, and so affects many important decisions, including capital requirements, merger approvals, and resolution processes for failed BHCs. Precision and accuracy of the NIC data are therefore important.

2.2.1 Data Quality

Ontologies can help maintain a dataset's internal coherence and integrity. Traditional metadata, such as SQL schemas, emphasize coherent identification (through primary key and uniqueness constraints), referential integrity (foreign key constraints), and domain integrity (restrictions on allowable types and values). However, traditional schemas and data dictionaries can exhibit errors, even for actively maintained databases like the NIC. For example, earlier versions of the public documentation [3] listed one `ENTITY_TYPE` code twice — `SLHC`, for *Savings and Loan Holding Company*. Another `ENTITY_TYPE` code — `IHC`, for *Intermediate Holding Company* – was missing altogether. An ontology does not guarantee the absence of gaps and inconsistencies, but it does facilitate automated scans of the metadata to help reveal them.

Other questions are more subtle. The NIC documentation [3] is ambiguous whether the meaning of "banking company" in the definition of the `EQUITY_IND` field in the ***Relationships*** table includes subsidiaries of type "financial holding company" (`ENTITY_TYPE=FHD` or `FHF`) as defined by the `ENTITY_TYPE` field. This ambiguity has ramifications, because `EQUITY_IND` is handled differently depending on the offspring type. If the offspring is a "banking company" (`EQUITY_IND=1`), the NIC records the exact ownership percentage in the `PCT_EQUITY` field. If the subsidiary is a "non-banking company" (`EQUITY_IND=2`), `PCT_EQUITY` indicates an approximate percentage as one of six brackets from the `PCT_EQUITY_BRACKET` field. An ontology can make precise these more subtle relationships and help enforce rules to improve the accuracy of the instance data.

2.2.2 Data Integration

Ontologies can assist in the integration of data from different sources. The alignment of entity identifiers from different datasets is an important practical challenge, as evidenced by the significant efforts of the LEI initiative [15]. Separately, the Federal Reserve maintains a mapping between RSSD ID identifiers and the PERMNO identifiers used by the dataset of the Center for Research in Securities Prices (CRSP) [11].

CRSP is an authoritative source for daily returns on U.S. equities and is widely used in research on investment portfolios. CRSP provides a largely orthogonal perspective on the state of large financial firms, so integrating these datasets is a powerful combination. While NIC gives details on BHCs' organizational structures, stock prices aggregate information from many investors on the value of firms' future cash flows. The PERMNO field identifies firms in the CRSP database. By aligning the RSSD ID identifiers for high-holder BHCs with the matching PERMNOs, the mapping [11] can augment the NIC with a rich daily time series of high-quality signals from the equity markets. Prior iterations of the Financial Entity Identification and Information Integration (FEIII) Challenge have addressed precisely this problem [20].

2.2.3 Semantic Modeling

For data structure, we downloaded the most recent data dictionary [3] and the NIC public dataset [9]. The next step is capturing authoritative definitions and explicit relations.

A semantic data model goes beyond rules describing the internal consistency of a dataset, to connect the data and metadata to external reference points in the wider world. For the NIC data, the most important points of connection are the authoritative definitions of key concepts in the U.S. Code [25] and Code of Federal Regulations [24]. Much of the value of the NIC data depends on a thorough understanding of this regulatory context. The NIC user community typically relies on accumulated human expertise to provide it. For example, it is likely not obvious to the uninitiated that the reification of *Intermediate Holding Company* (`ENTITY_TYPE=IHC`) is an artifact of the Fed's Regulation YY reporting requirements for foreign banking organizations. Because the legislative and regulatory corpus is highly structured, well documented, and authoritative, it is amenable to a formal semantic mapping to the NIC data and metadata. We have begun this semantic mapping process as part of this project.

3. Ontology Implementation

We chose the W3C standard Web Ontology Language (OWL) as our knowledge-representation format [27]. OWL's standard mechanisms for assigning intuitive semantic meaning give it advantages over other common data modeling languages, such as UML, XSD, and SQL. These traditional alternatives have limited tooling for describing semantic relationships based on domain knowledge, focusing instead on data integrity as defined by the internal consistency of the data themselves. The Fed's existing process relies on SQL schemas, data dictionaries, and expert curation. We are instead interested in applying a formal ontology to improve data quality and facilitate data integration by exposing expert domain knowledge in a human- and machine-readable way.

OWL's primary uses are flexible data modeling and efficient automated reasoning. We use OWL 2 to exploit the expressive power of direct semantics and description logics (DL) of efficient DL reasoners for consistency checking and inferencing [28]. By leveraging the OWL semantic data model and Resource Description Framework (RDF) instance data, a formal ontology can support simpler and more flexible queries than traditional SQL, for example, identifying all the explicit relations among financial institutions, such as owns, controls, advises, etc. that are only implicit in the original data tables. Improved clarity and coherence of the metadata could help reduce regulatory burden for reporting firms, while also improving the usefulness of the data to regulators. **Table 1** lists the ontology constituents in our framework, and **Figure 1** depicts their interrelationships.

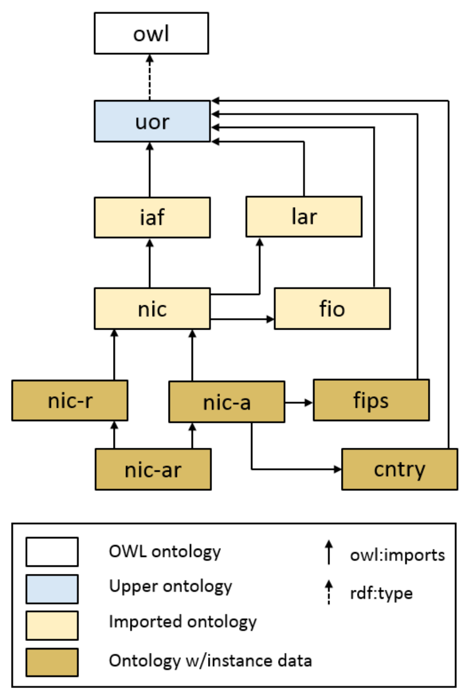
We use an incremental and iterative approach in developing the prototype ontologies. We use federated ontologies, instead of a monolithic enterprise ontology, to simplify ontology maintenance and target domain knowledge. The financial institutions ontology (FIO) is a taxonomy of legally defined financial institutions. The FIPS ontology covers the Federal Information Processing Standard. We apply requirements analysis in systems engineering and software engineering, starting with the identification of use cases and accompanying ontology competency questions to answer simple SPARQL queries against the sample instance data. Our primary use-case in the current iteration is to organize the domain knowledge and integrate relational data in a knowledge graph (NIC-AR) for queries and visualization. We developed a local upper ontology (UOR) to harmonize these federated ontologies. A competency question provides the terms and relations among them. We conferred with subject matter experts to validate the text interpretation and translation to OWL to ensure a robust ontological representation that can both meet current needs and be extensible for future requirements. In future implementation work, we will test the ontology's ability to address the use cases and answer competency questions with the sample data. The ontology provides semantic data models that are flexible for instance data to be searched and analyzed consistently.

Table 1. Ontology Constituents

Label	Description
OWL	Adherence to core standards for basic types
UOR	Local upper ontology
IAF	Information asset - financial ontology
LAR	Ontology of laws, acts, and regulations
FIO	Financial institutions ontology
NIC-R	NIC relationships, with instance data
NIC-A	NIC entity attributes, with instance data
NIC-AR	NIC attributes and relationships, merged
FIPS	Ontology of FIPS U.S. city and state codes
CNTRY	Ontology of country codes

Source: Authors' analysis

Figure 1. Structure of the Ontology Framework



Source: Authors' analysis

3.1 Modeling Ownership and Control

The NIC is a large database with many uses. We focus our attention on the ownership and control relationships among entities with a RSSD ID, including BHC subsidiaries.

The NIC *Relationships* table describes the connections between parent and offspring entities, with each uniquely identified by its RSSD ID (`ID_RSSD_OFFSPRING`, `ID_RSSD_PARENT`). The fields in this table describe the timing of relationships (`DT_START`, `D_DT_START`, `DT_END`, `D_DT_END`, `DT_RELN_EST`, `D_DT_RELN_EST`), their legal bases (`MB_COST`, `PCT_EQUITY_BRACKET`, `PCT_EQUITY`, `PCT_EQUITY_FORMAT`, `PCT_OTHER`, `OTHER_BASIS_IND`), reasons for creating and/or ending them (`REASON_ROW_CRTD`, `REASON_TERM_RELN`), and whether and how the relationship is regulated (`RELN_LVL`, `REGK_INV`, `REG_IND`). The fields are actually statements about the statements that parent entities are related to offspring entities. Meta knowledge about certain facts are attractive cases for RDF reification. We are considering testing before adopting the singleton property approach [21] with formal semantics. This approach is claimed to be simple and intuitive. An alternative approach is suggested in the RDF Primer [23], but this comes without formal semantics and not commonly used in practice.

3.1.1 Statistics

We use an incremental approach to import the NIC ontology and data instances sequentially into a knowledge graph (NIC-AR) in RDF, for data queries and visualization (see **Table 1**), ensuring each import is complete and correct. Our current knowledge graph has partial data from both the ***Attributes-Active*** and the ***Relationships*** tables. There are 81,371 instances of *Offspring Entity with RSSD ID*, forming 223,651 *directly holds* relationships, with 32,146 instances of *Parent Entity with RSSD ID*. Among these *Entity with RSSD ID* instances, 71,917 are *Active Entity with RSSD ID*.

3.1.2 Equivalence classes and logical definitions

The data dictionary lists 43 unique values of *Entity Code*, where each identifies a unique NIC `ENTITY_TYPE`.

> **COLUMN: ENTITY_TYPE**
> **DATA TYPE: CHARACTER [4]**
> **DESCRIPTION: ENTITY TYPE**
> The Entity Type field is derived from other fields. The entity types listed below are in alphabetic sequence by entity code.
> **AGB Agreement Corporation - Banking**
> **AGI Agreement Corporation - Investment**
> **BHC Bank Holding Company**
> **CPB Cooperative Bank …**

To represent accurately and unambiguously the true meaning in the NIC ontology, *Entity Code* is a subclass of *Identifier*, a class imported from the UOR ontology. UOR is a local upper ontology to harmonize domain ontologies for integration and navigation. We modeled *Entity Code* in a logical definition as a collection of 43 unique instances using `owl:equivalentClass`.

{AGB, AGI, BHC, CPB, DBR, DEO, DPS, EBR, EDB, EDI, FBH, FBK, FBO, FCU, FEO, FHD, FHF, FNC, FSB, IBK, IBR, IFB, IHC, INB, ISB, MTC, NAT, NMB, NTC, NYI, PST, REP, SAL, SBD, SCU, SLHC, SMB, SSB, TWG, UFA, UFB, USA, USB}

We connected each of these instances in the model via *is used to identify* to a subclass of *Entity Derived by NIC*. The latter is also modeled in a logical definition using `owl:equivalentClass`.

{*AgreementCorporation-Banking,*
AgreementCorporation-Investment,
BankHoldingCompany, … }

The FFIEC site provides a text definition for most subclasses of *Entity Derived by NIC*. In the NIC ontology, we documented a definition for each of these subclasses, with its source as a value of an annotation property *definition* imported from the UOR ontology. These semi-authoritative definitions are useful for reference and will be compared with those in the FIO ontology during mapping.

3.1.3 Property chains

In the NIC ontology, *holds* is an object property with two sub-properties, *directly holds* and *indirectly holds*. By using the `owl:propertyChainAxiom`, one can infer an *indirectly holds* relationship, to verify if available relationship data are accurate.

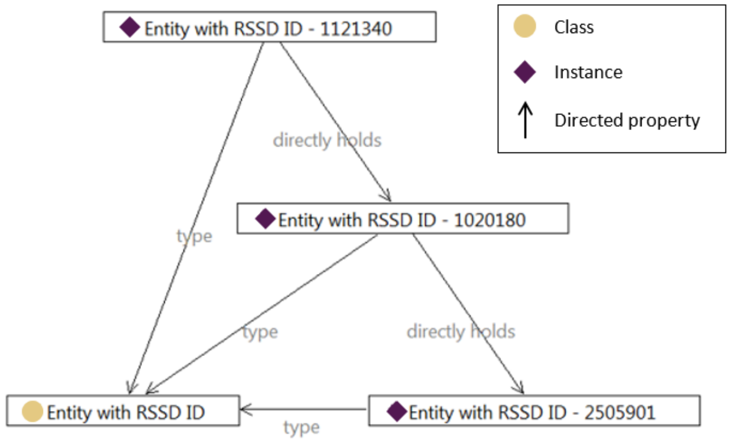
[*directlyHolds*, *directlyHolds*]
[*directlyHolds*, *indirectlyHolds*]
[*indirectlyHolds*, *directlyHolds*]
[*indirectlyHolds*, *indirectlyHolds*]

In addition, any of these pairs result in a new instance of the *indirectly holds* relationship. We will test them when the *indirectly holds* data are imported from the remaining ***Relationships*** table into this knowledge graph.

3.1.4 Reification

We avoid RDF reification for the *Reason for Termination of the Relationship* field (`REASON_TERM_RELN`) in two ways. First, the `nic:isTerminatedRelationshipBecauseOf` object property links `nic:OffspringEntityWithRSSDID` to instances of the class `nic:ReasonForTerminationOfRelationship` (there are six such instances in the ontology). This would work when the knowledge base is limited. A better alternative is to embed the reason in a finite number (six) of object sub-properties of `nic:terminatesRelationshipWith`, to allow direct links between the relationship pair, a parent entity and an offspring entity. This is a kind of reification and will work well when reasons are finite.

Figure 2. Ontology Provides Meaning to *Relationships* Data

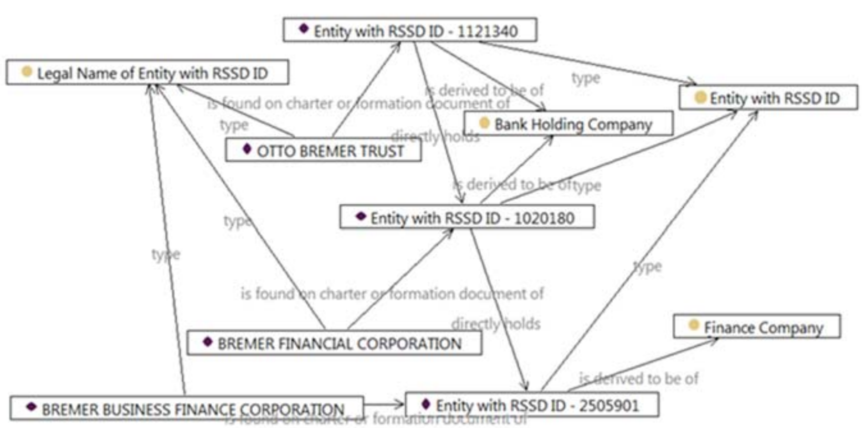


Sources: Federal Reserve Board's National Information Center, authors' analysis

An ontology consists of a set of precise descriptive statements in the form of subject - predicate – object triples. It describes the universal things in reality as classes and the explicit relations between them in an area of interest. In practice, ontologies organize the domain knowledge and integrate the relational data that may be in siloes, e.g., tables or datasets, into a knowledge graph. Ontologies can be shared as data models for applications or as descriptions of a domain area for new users. Our earlier Form PF Ontology project [8] demonstrates that financial ontologies can provide data meaning and facilitate search (as in **Figure 1** here), and can function as a semantic data model (as in **Figure 2**).

We have developed prototype formal ontologies to integrate data from ***Relationships*** and ***Attributes-Active***. In **Figure 2**, the ontology provides meaning to the data in the ***Relationships*** table. Three instances of *Entity with RSSD ID* relate to one another via a transitive object property *directly holds*, where this relation is not explicit according to the data in the table. **Figures 3** and **4** depict how ontologies integrate data from two tables, enabling a flexible data model in analysis to query and display data in a knowledge graph.

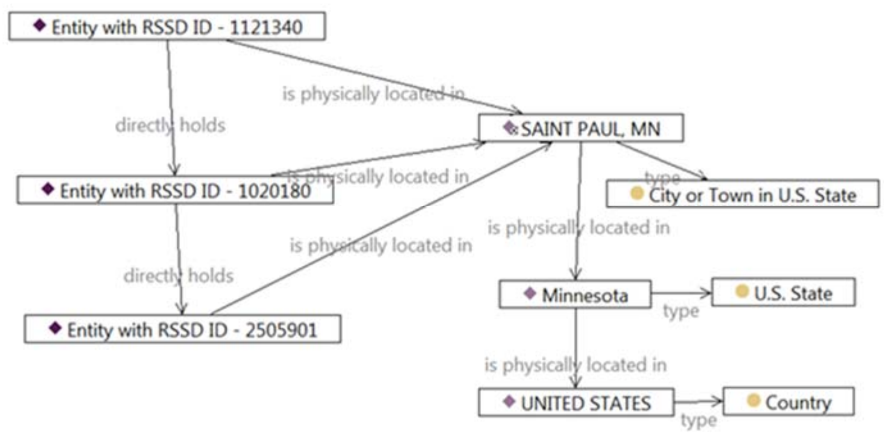
Figure 3. Flexible Data Models to Integrate Data in Silos



Sources: Federal Reserve Board's National Information Center, authors' analysis

Figure 3 depicts data integrated from the ***Relationships*** and ***Attributes-Active*** tables. Each of three instances of *Entity with RSSD ID* is connected to its own official name (*Legal Name of Entity with RSSD ID*) via an object property, *is found on charter or formation document of*. Each now also connects to an *Entity Type*; via another object property, *is derived to be of*, which is a derived relation described in text of the data dictionary [3], but not explicitly asserted in the ***Attributes-Active*** table.

Figure 4. Semantic Integration and Querying the Ontology



Sources: Federal Reserve Board's National Information Center, authors' analysis

Figure 4 shows how the same linked data (knowledge graph) can be queried flexibly and displayed graphically. The physical locations (city, state, and country) for each of three instances of *Entity with RSSD ID* are linked via a transitive object property, *is physically located in*. This relationship is transparent and intuitive here, but not explicit in the original data table.

4. Conclusions and Next Steps

We introduce a prototype NIC data ontology that intuitively organizes the domain knowledge (metadata) in a human- and machine-readable manner and integrates siloed relational data from two tables into a knowledge graph, which can be queried in flexible data models. Our next steps are to develop fully the NIC data ontology to integrate all data in the NIC public data tables for applications. We will consider developing additional harmonized ontologies to extend the domain knowledge of other regulatory datasets that the Office of Financial Research has collected on behalf of the Financial Stability Oversight Council. The integrated regulatory data in a knowledge graph will serve as a powerful tool in monitoring financial stability and decision making. Flexible data models and visualizations will add value to the integrated datasets. In addition, ontologies may stand alone as a service for educational purposes to aid public understanding.

REFERENCES

[1] Berle, A., and Means, G. 1932. *The Modern Corporation and Private Property*. Transaction Publishers.

[2] Board of Governors of the Federal Reserve System. 2007. RSSD Data Dictionary, technical report. (June).

[3] ______. 2016. "Bulk Data Download: Data Dictionary and Reference Guide, Document Version 1.1," technical report (June). www.ffiec.gov/nicpubweb/content/DataDownload/NPW%20Data%20Dictionary.pdf.

[4] ______. 2016. "Instructions for Preparation of Report of Changes in Organizational Structure, Reporting Form FR Y-10," technical report (Oct.). www.federalreserve.gov/reportforms/forms/FR_Y-1020180228_i.pdf.

[5] Board of Governors of the Federal Reserve System and Federal Deposit Insurance Corporation. 2016. "Resolution Plan Assessment Framework and Firm Determinations," technical report (April). www.federalreserve.gov/newsevents/press/bcreg/bcreg20160413a2.pdf.

[6] Cetorelli, N. 1999. "Competitive Analysis in Banking: Appraisal of the Methodologies." *Economic Perspectives* (Federal Reserve Bank of Chicago) 23, 1 (Feb), 2-15.

[7] Fama, E., and Jensen, M. 1983. "Separation of Ownership and Control," *Journal of Law and Economics*, 26, 2 (June), 301-325.

[8] Fan, L., and Flood, M. 2016. "An Ontology of Form PF," *DSMM '16: Proceedings of the International Workshop on Data Science for Macro-Modeling, 2016 ACM SIGMOD Conference* (July). https://dx.doi.org/10.1145/2951894.2951901.

[9] Federal Financial Institutions Examination Council. 2018. "National Information Center: Bulk Data Download." www.ffiec.gov/nicpubweb/nicweb/DataDownload.aspx

[10] ______. 2018. "Central Data Repository (CDR) Glossary." https://cdr.ffiec.gov/cdr/public/cdrhelp/Glossary091505.htm.

[11] Federal Reserve Bank of New York. 2017. "Banking Research Datasets." www.newyorkfed.org/research/banking_research/datasets.html.

[12] Financial Stability Board. 2017. "2017 list of global systemically important banks (G-SIBs)," technical report www.fsb.org/wp-content/uploads/P211117-1.pdf.

[13] Flood, M., Kenett, D., Lumsdaine, R., and Simon J. 2018. "The Complexity of Bank Holding Companies: A New Measurement Approach." Office of Financial Research Working Paper Series no. 17-03, (March).

[14] Flood, M., Kwan, S., and Leonova, I. 2014. "Data for Microprudential Supervision of U.S. Banks," *Handbook of Financial Data and Risk Information, Vol. I*, Cambridge University Press, 426-489.

[15] Global Legal Entity Identifier Foundation and McKinsey & Co. 2017. "The Legal Entity Identifier: The Value of the Unique Counterparty ID," technical report. (Oct.). www.gleif.org/en/about-lei/mckinsey-company-and-gleif-creating-business-value-with-the-lei/.

[16] International Organization for Standardization (ISO). 2012. "ISO 17442:2012, Financial Services – Legal Entity Identifier (LEI)," technical report. (June).

[17] ______. 2015. "ISO/IEC/IEEE 15288:2015 – Systems and Software Engineering -- System Life Cycle Processes," technical report. www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=63711.

[18] Jensen, M., and Meckling, W. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics* 3, 4 (Oct.) 305–360.

[19] Legal Entity Identifier Regulatory Oversight Committee. 2015. "Statement on Individuals Acting in a Business Capacity," technical report (Sep.). www.leiroc.org/publications/gls/lou_20150930-1.pdf.

[20] National Institute for Standards and Technology. 2017. "FEIII: Financial Entity Identification and Information Integration." https://ir.nist.gov/dsfin/2017-challenge.html.

[21] Nguyen, V., Bodenreider, O., and Sheth, A. 2014. "Don't Like RDF Reification? Making Statements about Statements Using Singleton Property," *Proceedings of the International World Wide Web Conference.* https://dx.doi.org/10.1145%2F2566486.2567973.

[22] OECD Steering Group on Corporate Governance. 2007. "Lack of Proportionality between Ownership and Control: Overview and Issues for Discussion," technical report (Dec.). www.oecd.org/daf/ca/40038351.pdf.

[23] Schreiber, G., and Raimond, Y. (eds.). 2014. "RDF 1.1 Primer," *W3C Working Group Note* (June). www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/

[24] U.S. Government Printing Office. 2018. Electronic Code of Federal Regulations. https://www.ecfr.gov/.

[25] ______. 2018. United States Code. www.gpo.gov/fdsys/browse/collectionUScode.action.

[26] Vitali, S., Glattfelder, J., and Battiston, S. 2011. "The Network of Global Corporate Control," *PLoS ONE* 6, 10, e25995.

[27] W3C OWL Working Group. 2012. "OWL 2 Web Ontology Language, Document Overview, 2nd Ed. W3C Recommendations," (Dec.). www.w3.org/TR/owl2-overview/.

[28] ______. 2012. "OWL 2 Web Ontology Language, Direct Semantics, 2nd Ed. W3C Recommendations," (Dec.) www.w3.org/TR/owl2-direct-semantics/.