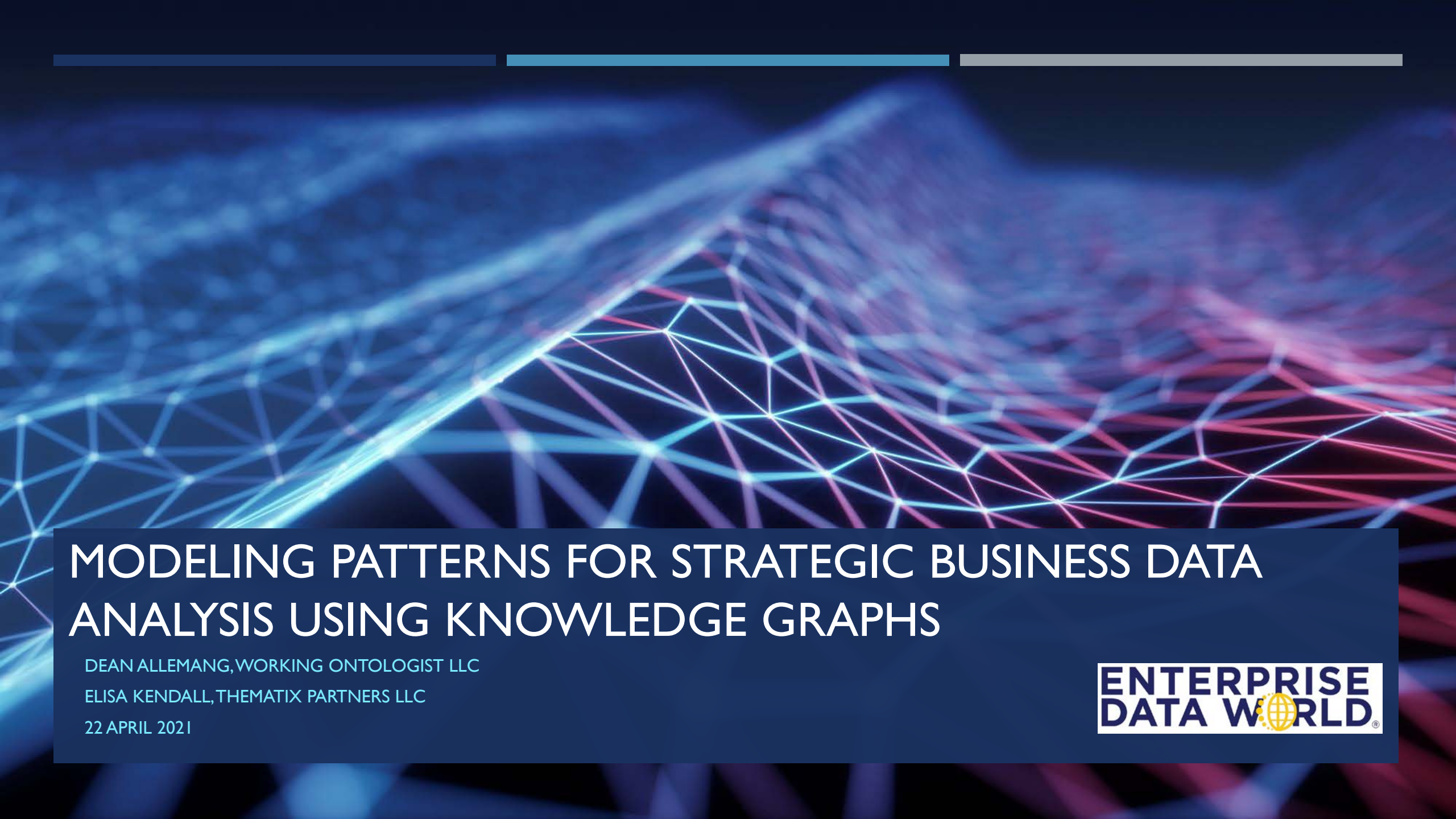


MODELING PATTERNS FOR STRATEGIC BUSINESS DATA ANALYSIS USING KNOWLEDGE GRAPHS

DEAN ALLEMANG, WORKING ONTOLOGIST LLC

ELISA KENDALL, THEMATIX PARTNERS LLC

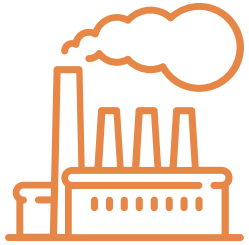
22 APRIL 2021

ENTERPRISE DATA WORLD®

SCOPES OF DATA MANAGEMENT

Application Data

Enterprise Data



Industrial Data

- **Regulatory data**
- **Product research**
- **Basic science**
- **Liability**
- **Market history**

WHY DO WE WANT TO MANAGE INDUSTRY DATA?

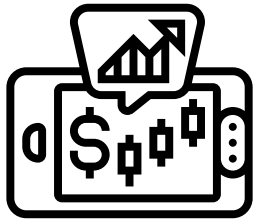
- Regulation
  - Regulators need to be able to specify what activity has to be reported
  - Regulators need to be able to understand the report
- Customer-centric data
  - Migrate data from one company to another (e.g., patients in medical centers)
  - GDPR, CCPA, personal rights to data
- Re-usable resources for Enterprise data
  - Code lists
  - Data models

HOW DOES A KNOWLEDGE GRAPH SUPPORT INDUSTRY DATA?

- Ontologies provide reference structure to data
  - "an address includes a country, subdivision and a postcode"
  - "an Account is held by an AccountHolder; a CustomerAccount is held by a Customer"

- Controlled vocabularies
  - Agrovoc
  - LCC (Languages, Countries and Codes)
  - NAICS / SIC (industry codes)
- Data sets
  - GLEIF
  - CHEBI

HOW TO USE AN INDUSTRY ONTOLOGY – SAMPLE USAGE SCENARIO



Company A is a stockbroker.
They keep records of transactions

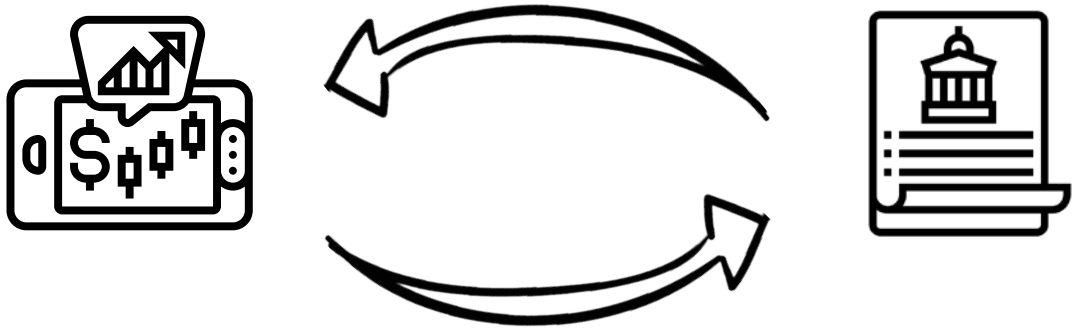
Purchase Order	Buyer	Stock	Volume	Total Price	Date
001	ACME	EXU	100	$2174	01/01/2021
002	FUJI	PBQ	50	$145	03/14/2021

From this, we can conclude that ACME owns EXU and FUJI owns PBQ. But that's not how it is represented.

Gov't Agency B
Keeps track of ownerships, for tracking who owes tax, supporting money laundering and fraud investigations, etc.

{"type":"Legal Entity",
"id": "ACME",
"owns":"EXU" … }
{"type":"Legal Entity",
"id": "FUJI",
"owns":"PBQ" … }

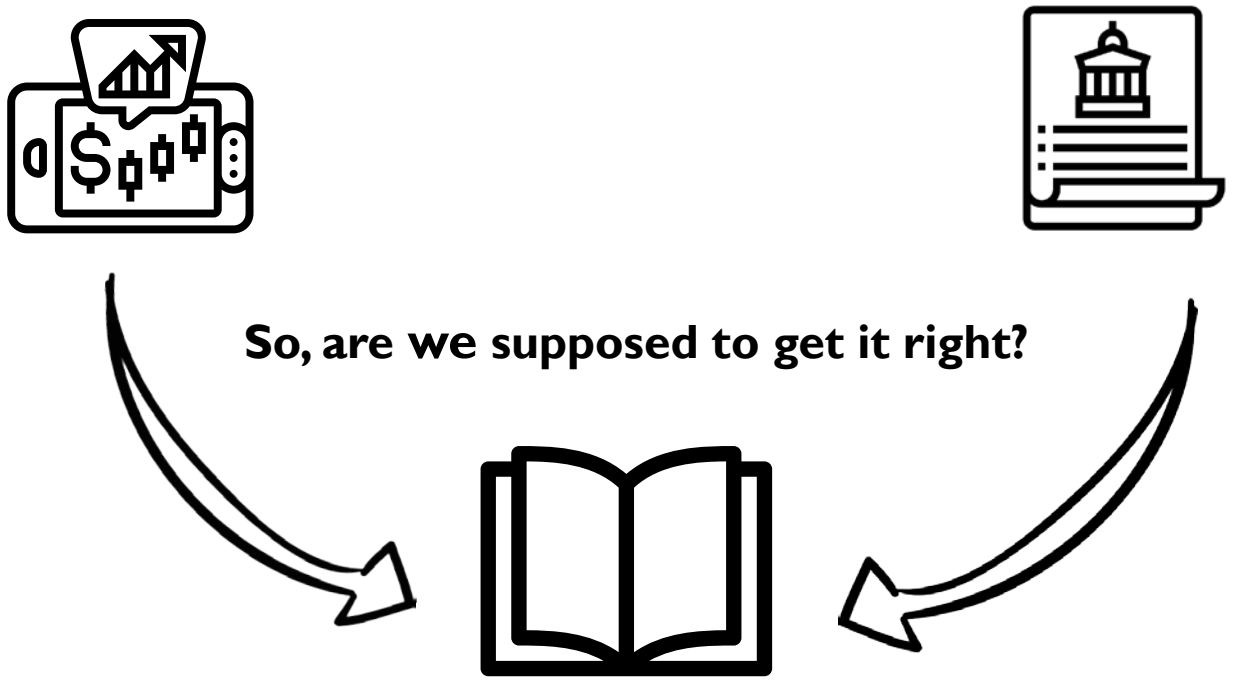
HOW TO USE AN INDUSTRY ONTOLOGY – SAMPLE USAGE SCENARIO



Which one is "right"?

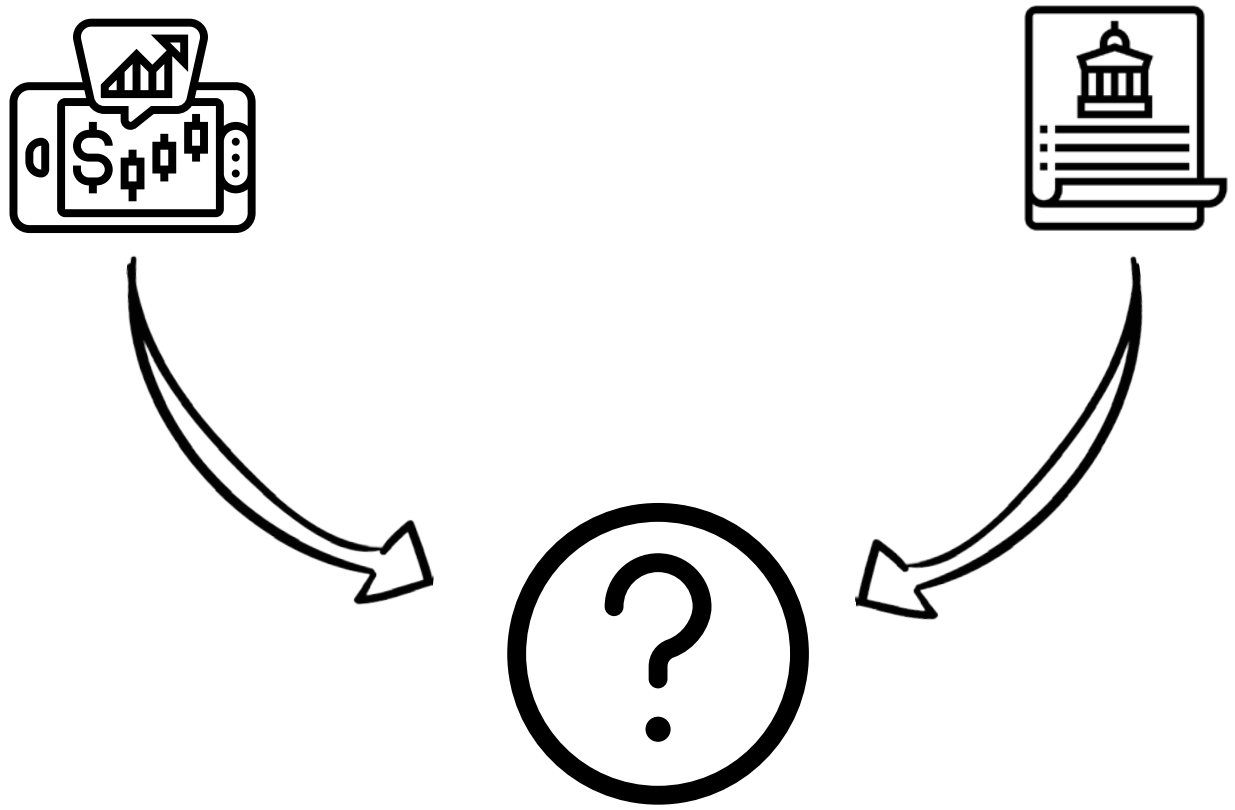
Neither one is "right"; each is fit for purpose!

HOW TO USE AN INDUSTRY ONTOLOGY – SAMPLE USAGE SCENARIO

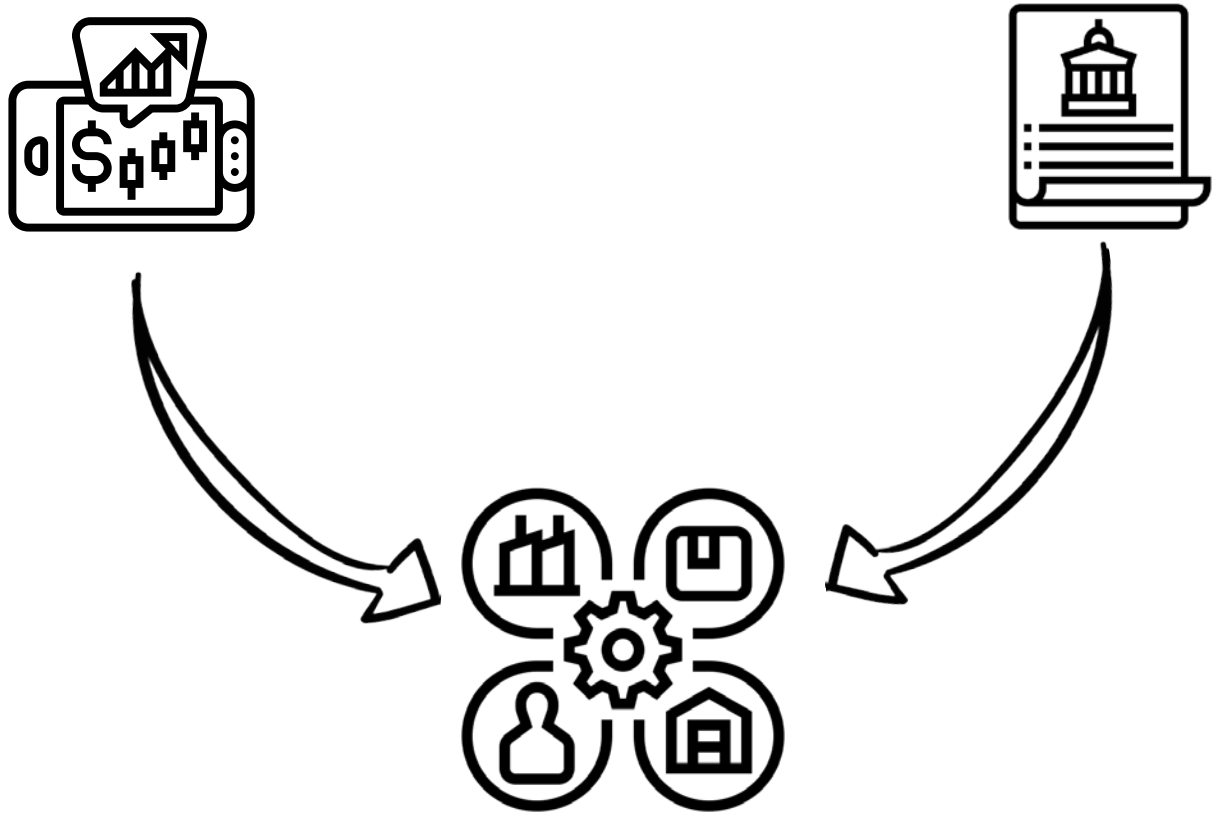


I'm sure that will make us popular.

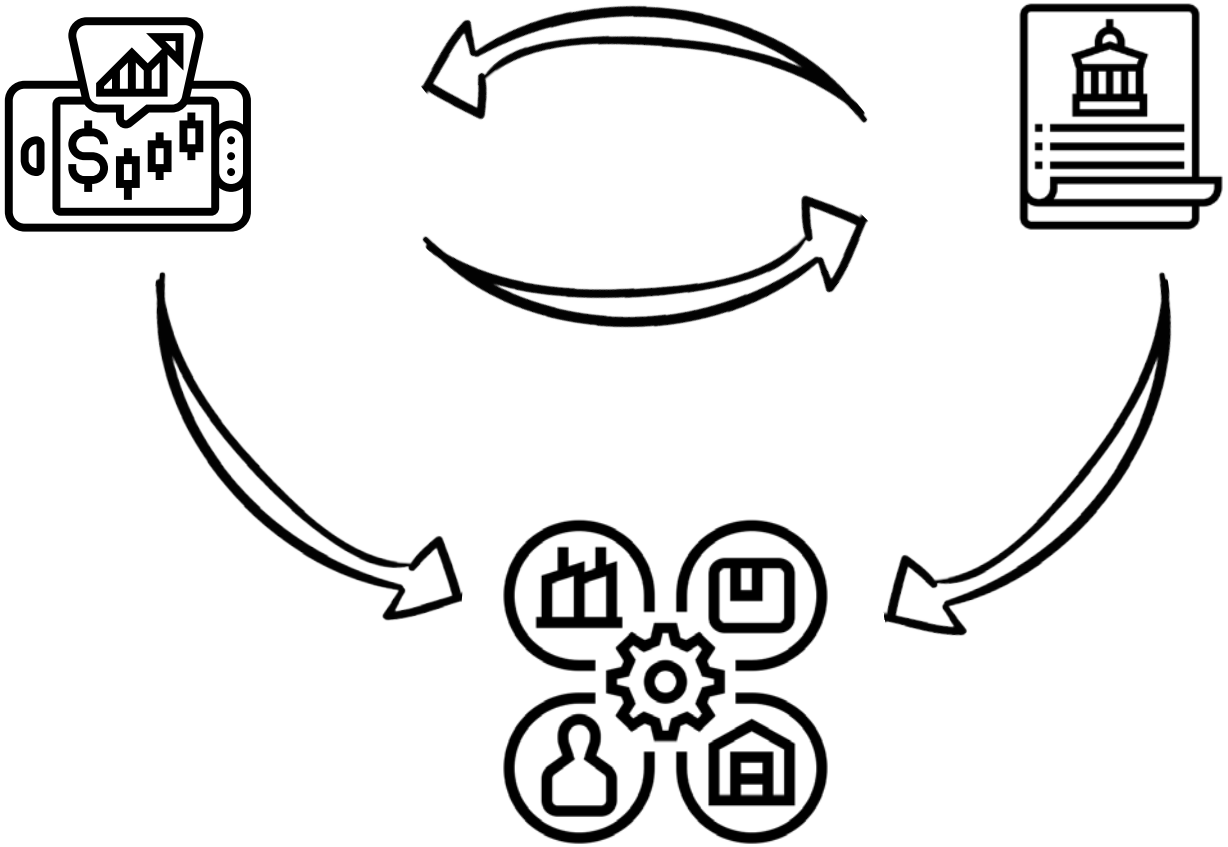
SO WHAT CAN WE DO TO HELP OUT THE INDUSTRY?



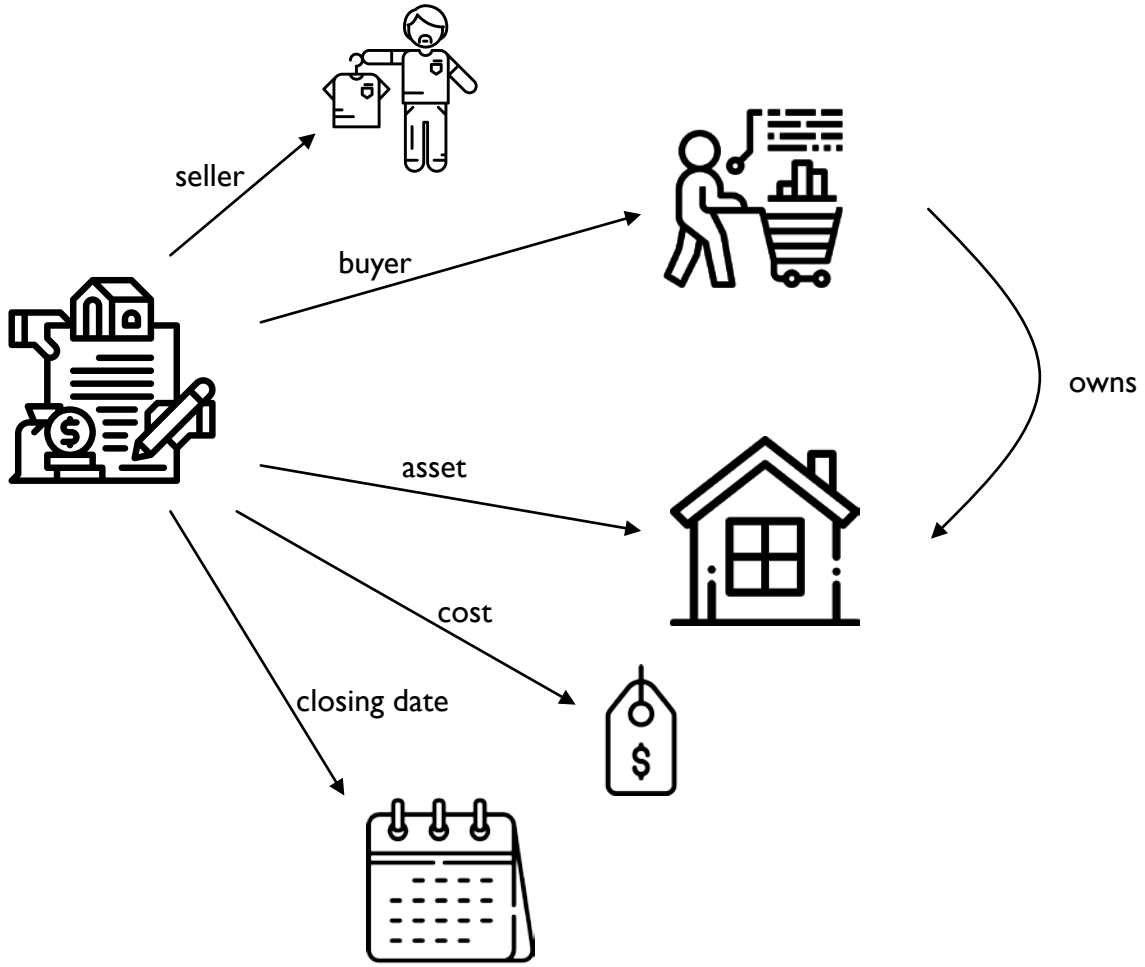
REFERENCE MODEL THAT CAN MAP TO EACH SOURCE MODEL …



… AND SHOW HOW THEY RELATE TO ONE ANOTHER



… AND SHOW HOW THEY RELATE TO ONE ANOTHER



WHAT IS AN INDUSTRY MODEL?

What is our job when we build an industry model?

Model the real world as it is?	No – each stakeholder has their own idea of what the 'real world' is; we are in no position to tell some of them that they are wrong.
Make a 'correct' model by fiat?	No – the stakeholders are not going to change what they are doing because we tell them to do it another way.
Make a model of the "common core" that everyone agrees on	Sort of – but this is pretty small. We'd like to do better …

WHAT IS OUR JOB?

- Anticipate what the models in the industry will look like
- Combine features we see in the anticipated model
- Design a model that will cover all of these features
- Evaluate the model against real world situations

ABOUT FIBO

- The Financial Industry Business Ontology (FIBO) is an industry-level ontology that provides standard terminology, relationships, and logic designed to help reconcile disparate language defining financial instruments and related knowledge
- Initially developed by data management professionals from a variety of institutions, led by the EDM Council in response to the 2008/2009 crisis and subsequent regulation in the EU and US
- Transformed to RDF/OWL in 2013, with increasingly robust development processes and governance
- First release as a joint Object Management Group (OMG) and EDM Council international standard in 2015
- Quarterly releases, developed by domain experts with guidance by professional ontologists, are published on the EDM Council site, freely available at https://spec.edmcouncil.org/fibo/, and available in GitHub at https://github.com/edmcouncil/fibo
- Work to publish a new baseline at OMG, is underway, planned for later this year

EDMCouncil

OMG OBJECT MANAGEMENT GROUP®

FIBO Financial Industry Business Ontology

USE CASES DRIVING FIBO DEVELOPMENT

- Original FIBO use case: develop an industry glossary that financial institutions and others could use to meet regulatory requirements – Dodd/Frank in the US and the MiFID II framework in the EU for regulating financial markets
- Additional requirements for data governance, data management, and enterprise glossaries were mandated in the EU by the Basel Committee on Banking Supervision (BCBS) for risk data aggregation and reporting
- Regulators in the US and around the world are demanding even more transparency and considering additional regulations in light of recent challenges related to
  - crypto currencies
  - hedge fund short selling and the Robinhood / GameStop debacle earlier this year
  - more recently, the Archegos Capital implosion and its reverberations through international markets
- Establishing a common basis for understanding these instruments, their behaviors, who the counterparties are and how broadly their exposure extends and related risk mitigation strategies remains critical
- The Financial Transparency Act (H.R. 4476) calls for ontologies to establish consistency in data definitions, formats and content across financial regulators to facilitate information sharing to help identify issues, and is likely to be extended to the commercial sector once it is enacted

WHAT IS A PATTERN?

- Reusable arrangement of nodes and edges in a graph
- Reusable arrangement of classes, relationships, individuals and logical axioms in a knowledge graph

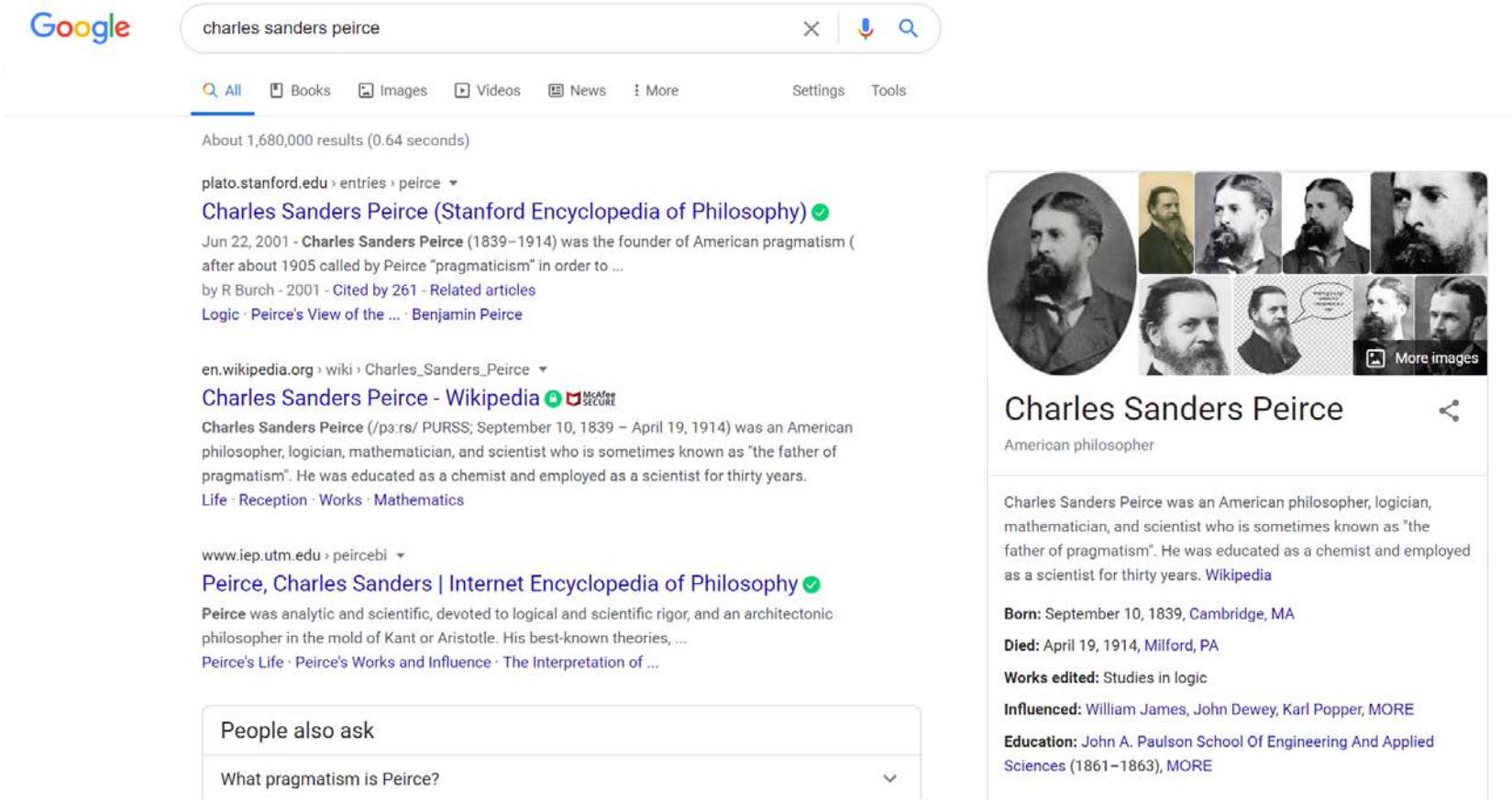
- Patterns help manage complexity while enabling the variation that is common in industry-level ontologies
- Many common patterns are recognizable – data modelers are familiar with party – role relations, for example
- Using a library of patterns can increase quality, improve consistency and interoperability, and reduce the learning curve

"SIMPLE" PATTERNS

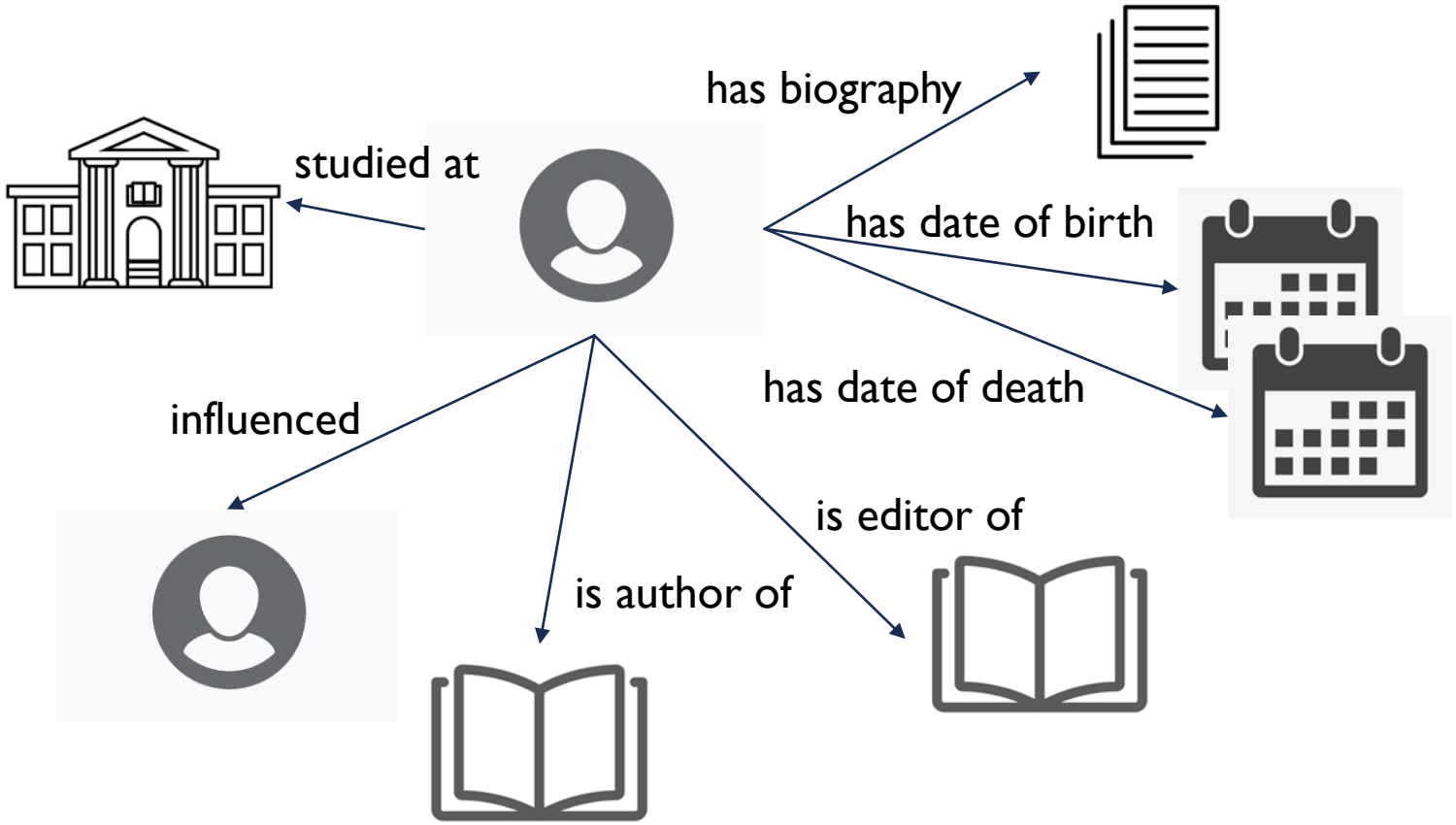
- Things vs. Strings – representing structured information about things in a graph rather than simply using string-valued attributes that limit reusability and interoperability
- Assigning names to things – name resolution across large corporations – for any 360º view of your customer in a large enterprise, regulations in banking known as 'know your customer' (KYC), etc.
- Identifying things – identity resolution typically more challenging than name resolution
- Relating codes to things – codes assigned to industries, markets, places, and many other business elements, similar pattern to identifiers
- Relating classification schemes / classifiers to things, such as asset classes for instruments, for which data providers and regulators have their own, conflicting schemes, again follow a similar pattern to identifiers

THINGS VS. STRINGS

- Google popularized the term "things vs. strings" in 2012 when they established their knowledge graph group internally, with the goal of providing structured information about people, places, events, and other "things" in response to user searches rather than lists of web links containing words matching the search terms



THINGS VS. STRINGS



Charles Sanders Peirce

American philosopher

Charles Sanders Peirce was an American philosopher, logician, mathematician, and scientist who is sometimes known as "the father of pragmatism". He was educated as a chemist and employed as a scientist for thirty years. Wikipedia

Born: September 10, 1839, Cambridge, MA

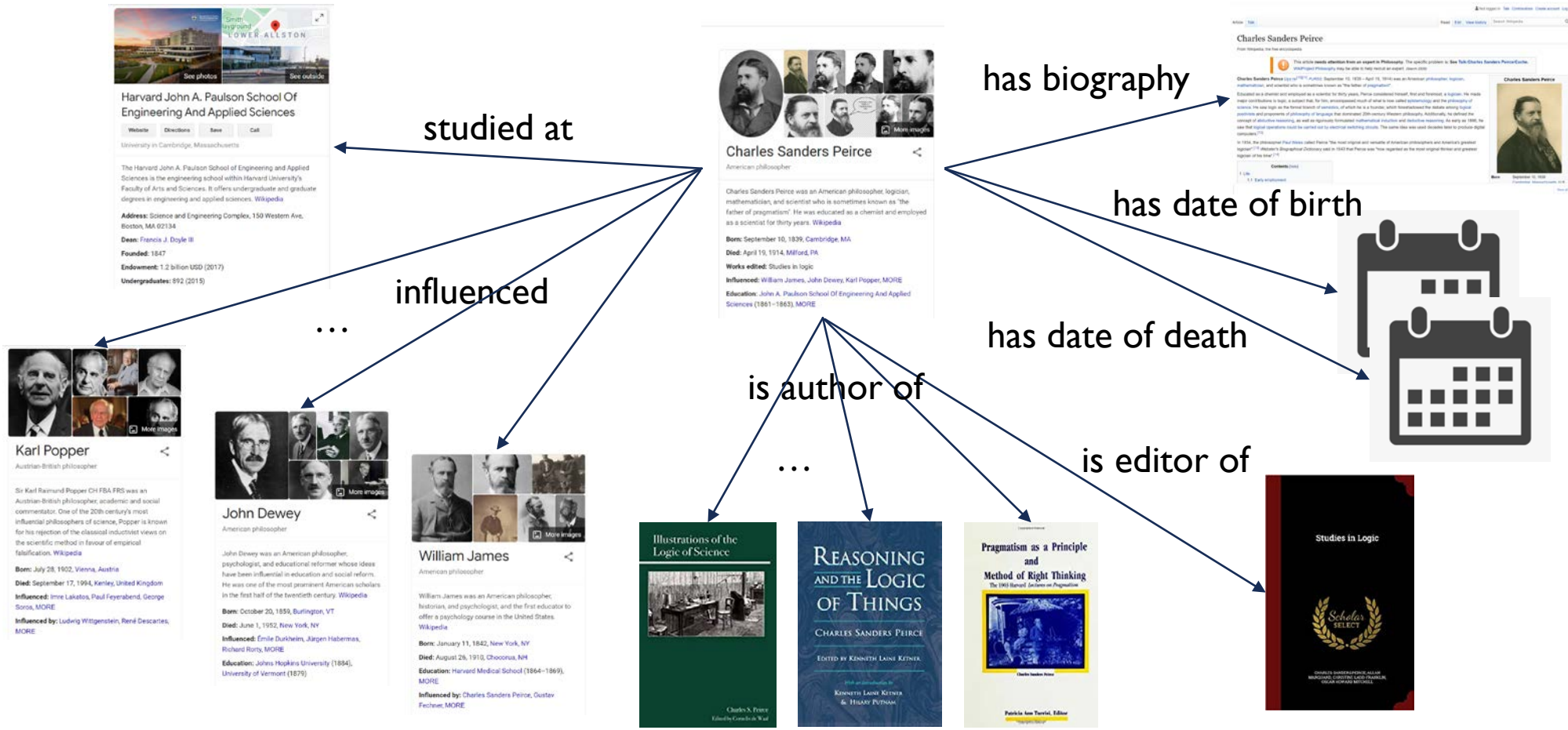
Died: April 19, 1914, Milford, PA

Works edited: Studies in logic

Influenced: William James, John Dewey, Karl Popper, MORE

Education: John A. Paulson School Of Engineering And Applied Sciences (1861–1863), MORE

THINGS VS. STRINGS

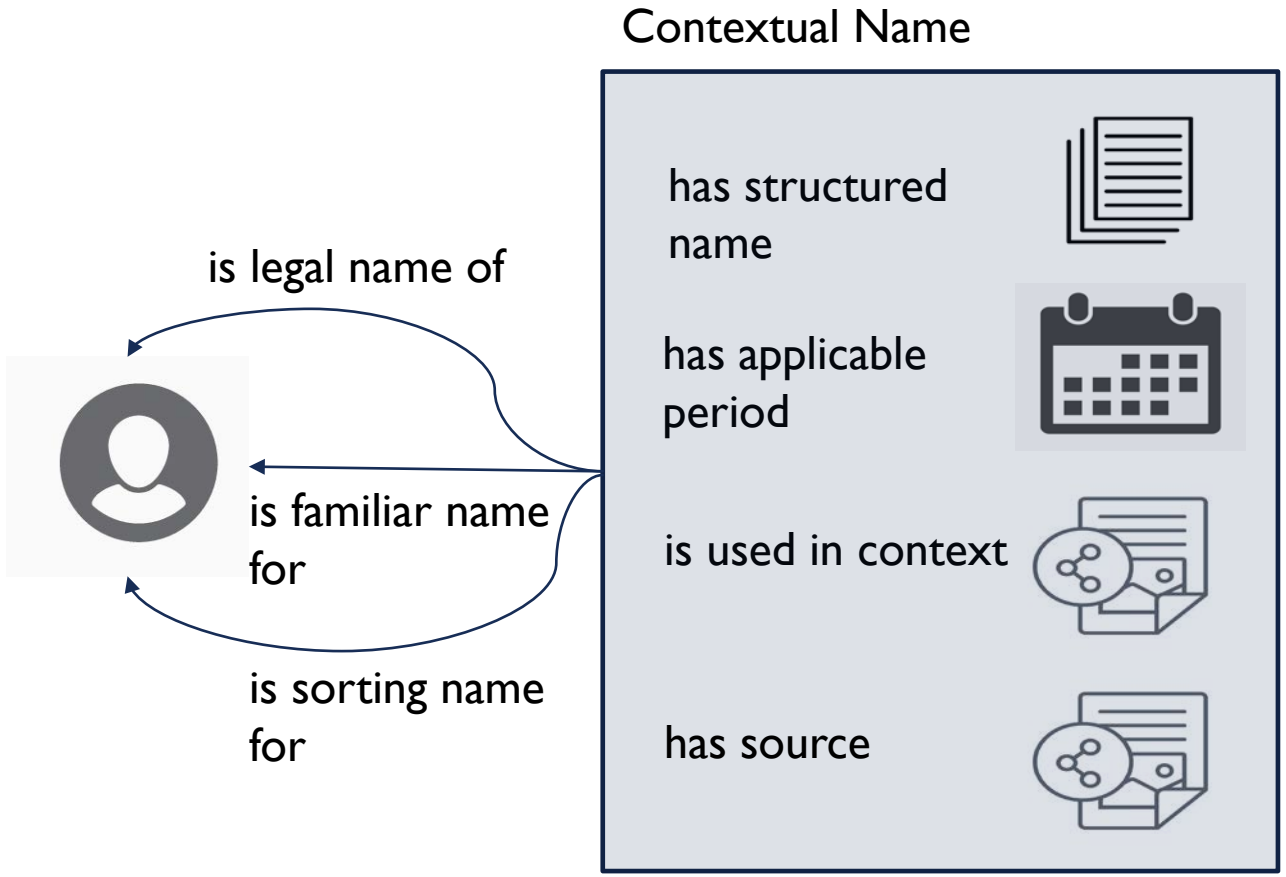


NAMING PATTERNS

- Element hasName string … or
- Structured names – names that themselves have elements
  - apply or became known as of some date
  - have provenance
  - are used/valid under certain circumstances
- Prior name, full legal name, familiar name (nickname) – these are separate structural names
- Digital footprint includes a structural name and other features that together aid in identity resolution

Legacy Person Name ?	But … in today's world …
Prefix	preferred under what conditions, honorific or something else?
Given / First Name	does everyone have one?
Middle Name	repeated?
Middle Initial	do we need an option for either?
Surname / Last Name	hyphenated, multiple? as of when? cultural variations? does everyone have one?
Suffix	repeated?
Prior Name	name on birth certificate, married name, changed name? …
Full Legal Name	as of when?
Nickname	preferred familiar name? / shortened name under what conditions?

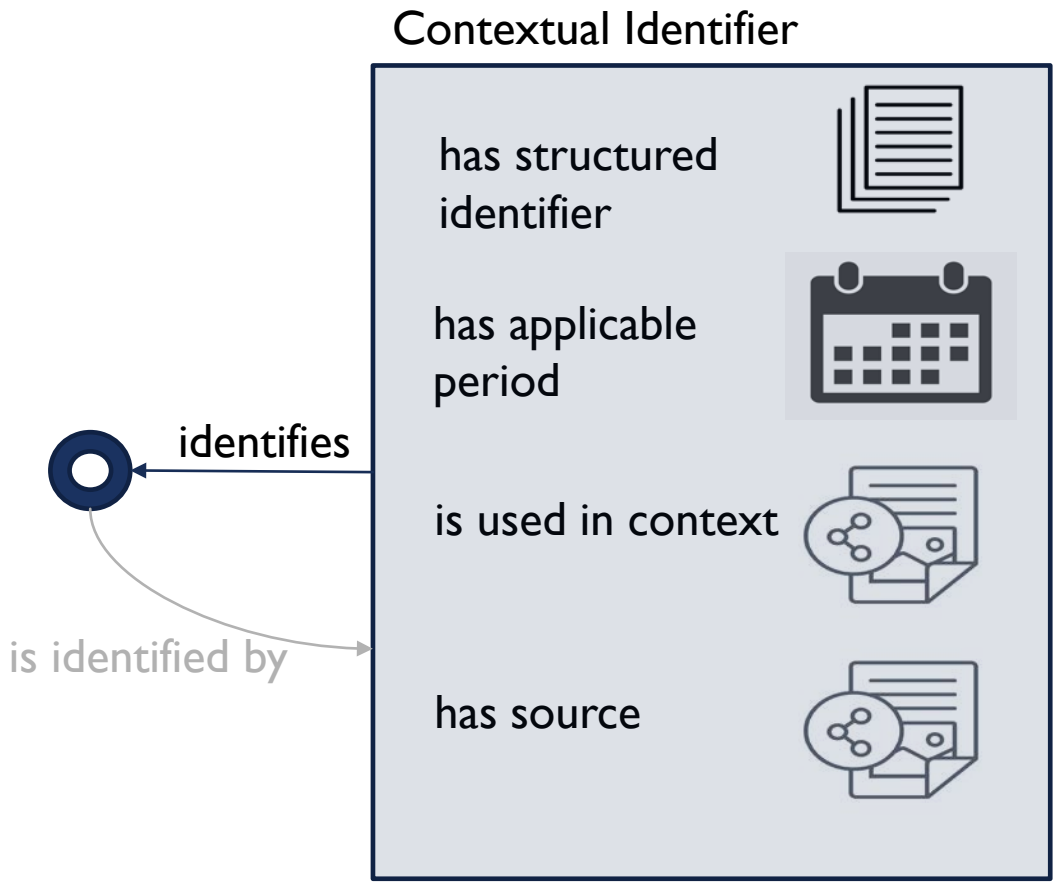
NAMING PATTERNS



- The structure for a person's name should include a full name as a single value, a family name for sorting, and prefix/suffix
- Use a similar pattern for organizations
- Note that the relationship between a contextual name and the person point ***from the name to the individual***, not the other way around
- Inverse relations (has legal name, has familiar name, has family name, has prior name, has sorting name etc.) can be used to infer that a particular name is associated with a person
- Facilitates mapping to various resources for diverse applications
- Supports normalization across resources, contexts and reusability in general

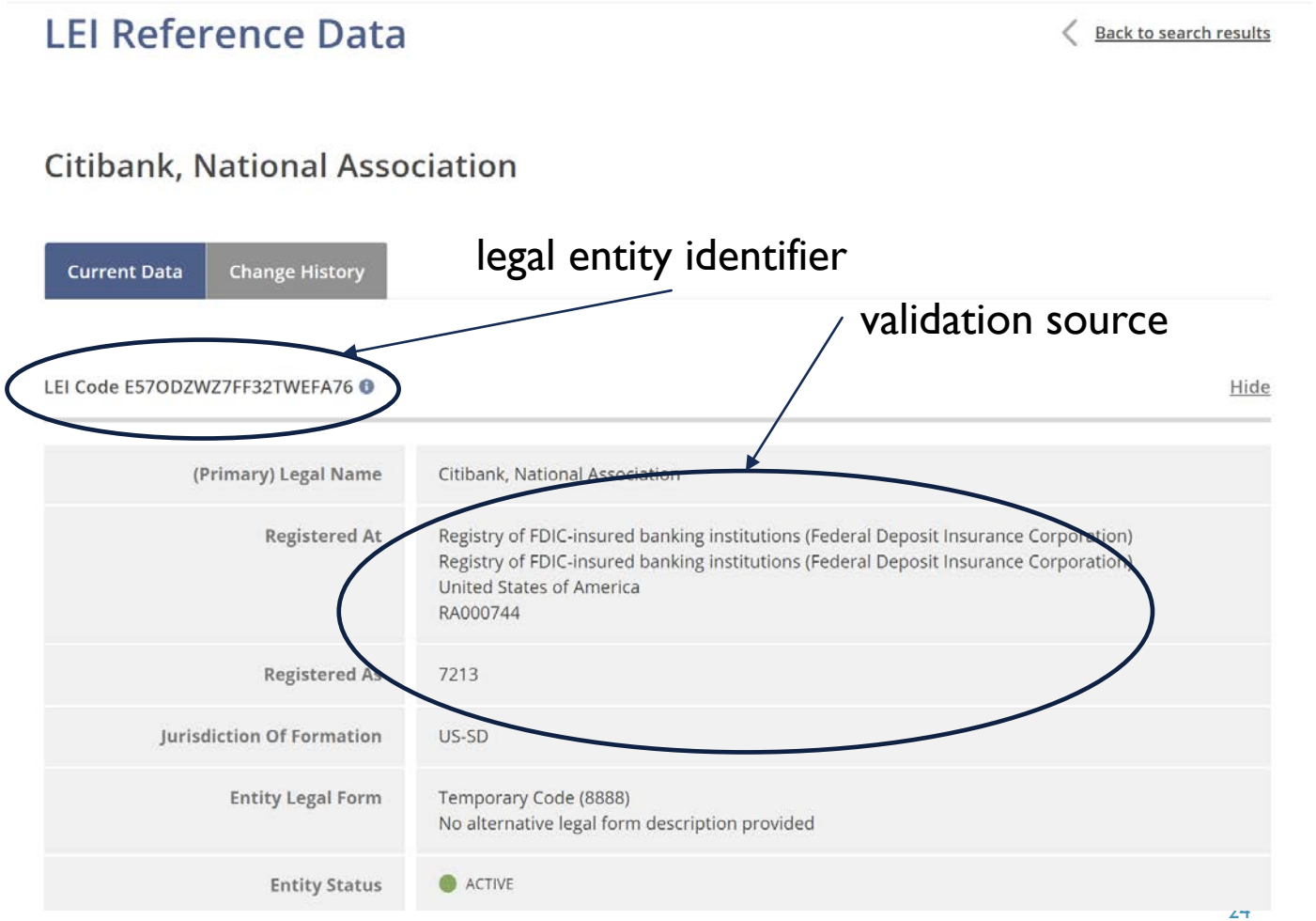
IDENTIFIERS & IDENTIFICATION PATTERNS

- An ***identifier*** identifies something, not the other way around – use an inverse, (is identified by), to infer what it identifies
- Structured identifiers – including identifiers that are composites, e.g., ISIN (with country code) vs. NSIN (without country code)
  - are registered with some registration authority
  - are defined in some sort of scheme
- Identifiers are things themselves, not strings

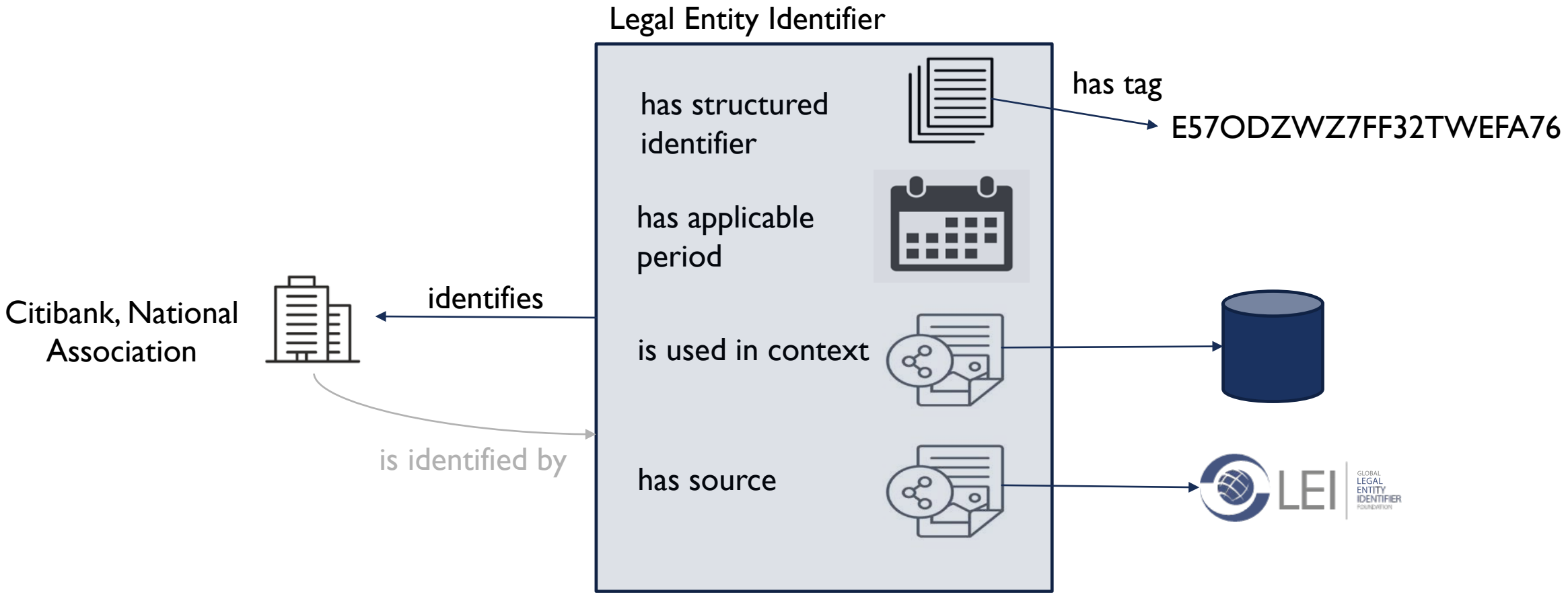


IDENTIFIER EXAMPLE

- Citibank, National Association is identified by a Legal Entity Identifier (LEI)
- Banks and regulators require LEIs to identify counterparties to financial contracts
- Search is powered by ontologies derived from FIBO and enhanced to support the Global LEI Foundation (GLEIF) process
- Data for 1.6+ M records is available on data.world using the GLEIF ontologies



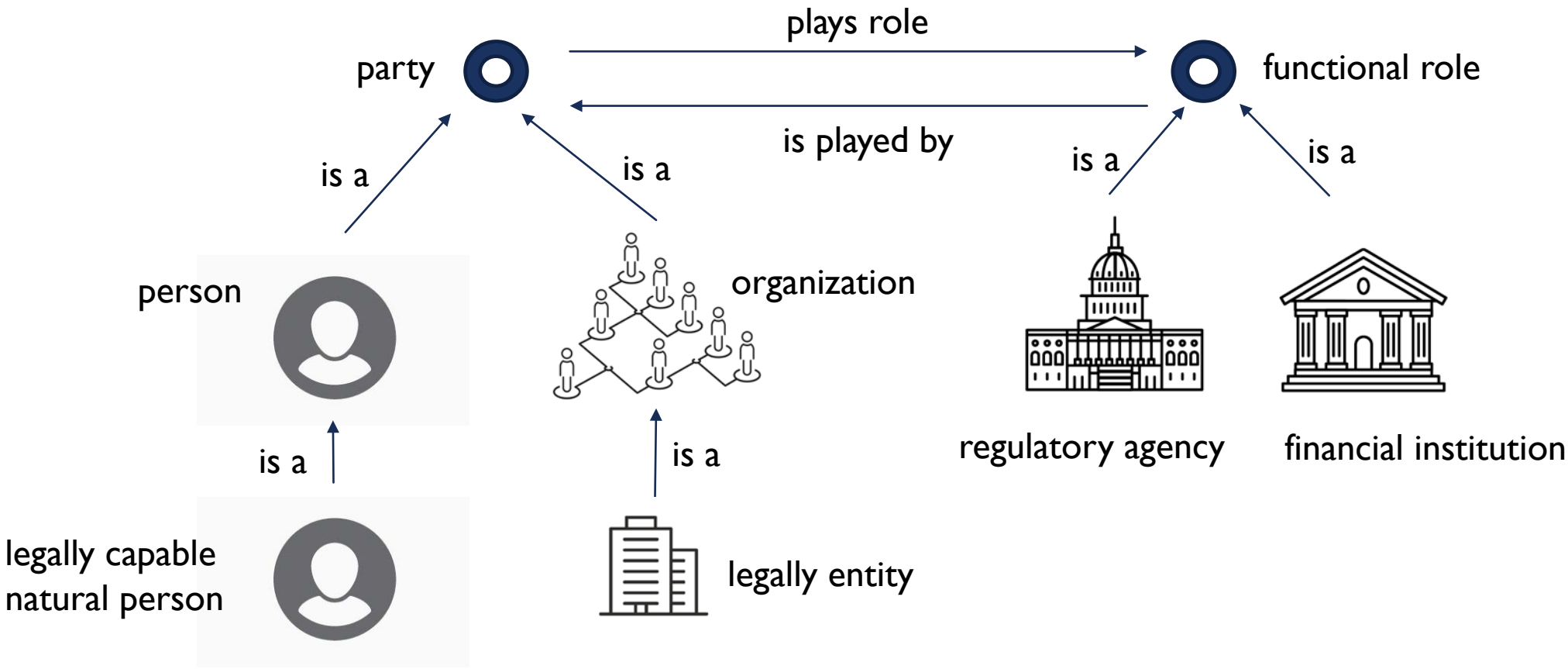
IDENTIFIER EXAMPLE



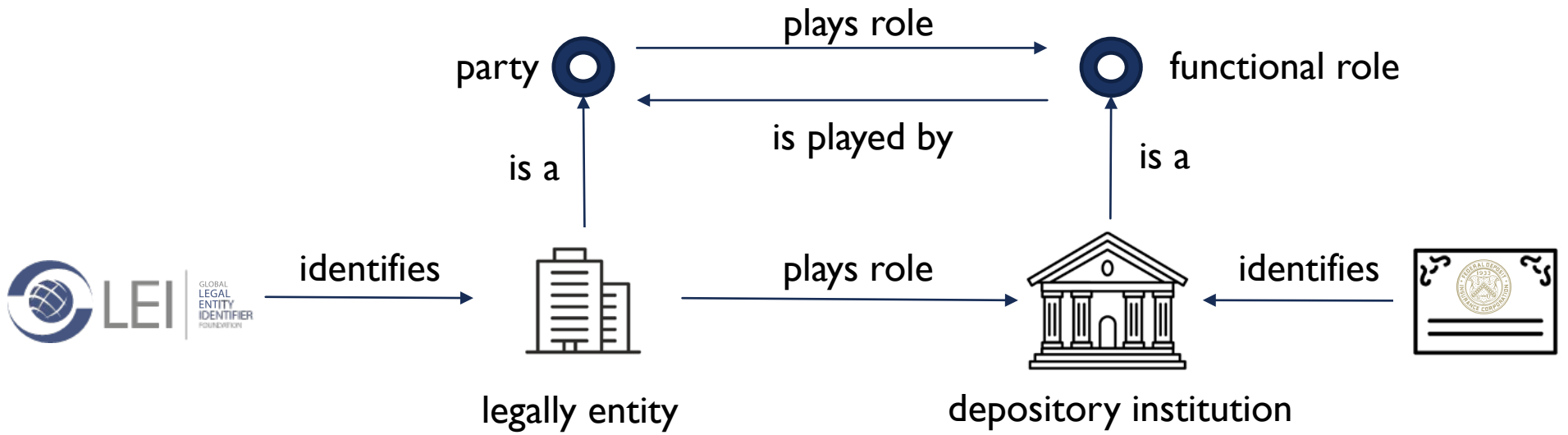
"MORE COMPLICATED" PATTERNS

- Associating parties (people and organizations) with the roles they play
- Identifiers reprised – when does an identifier apply to a party or thing vs. a role
- Situational analysis – linking parties and roles to situational patterns that are time bound (*e.g.*, ownership, control, authorization, membership, employment …)

PARTIES VS. ROLES



IDENTIFIERS FOR PARTIES OR ROLES?

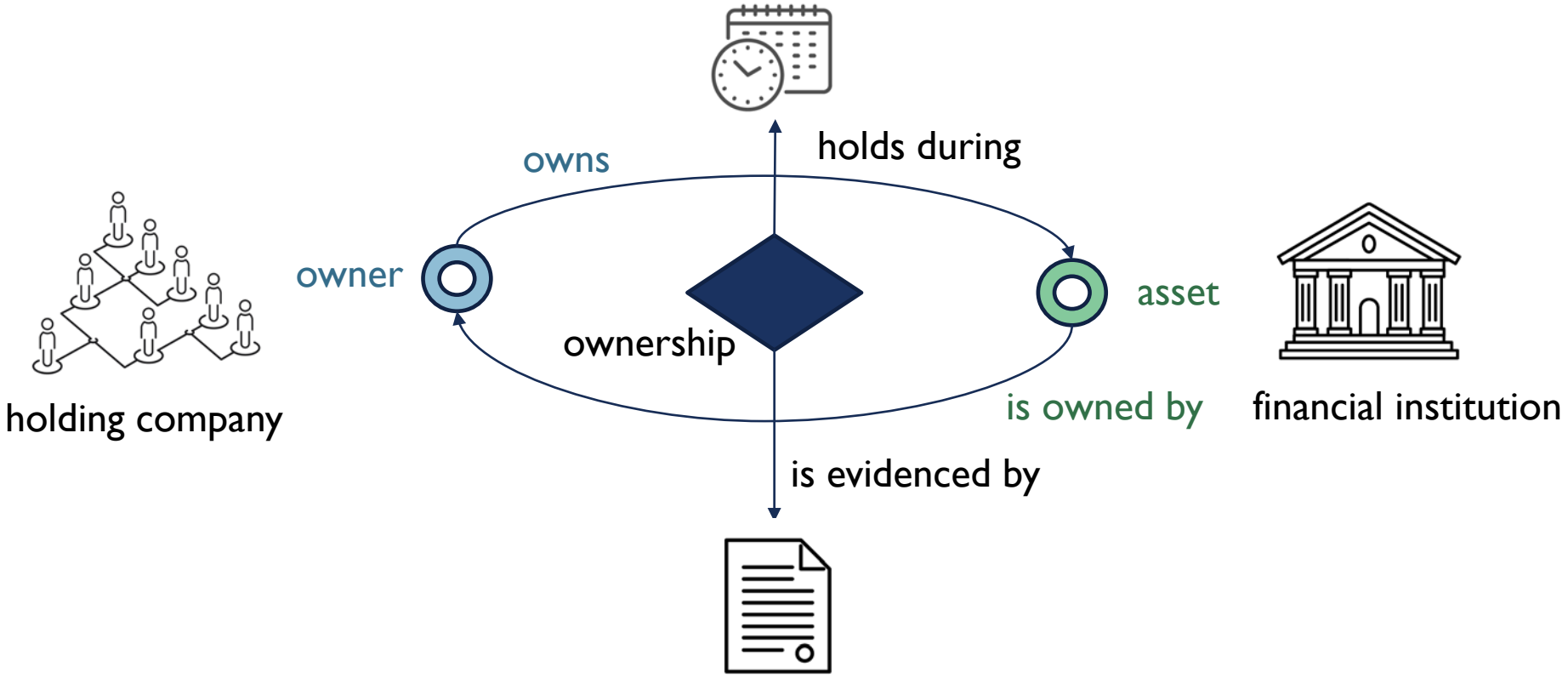


- LEIs, business identifiers issued by state governments, and others identify legal entities directly
- FDIC certificates, Routing Transit Numbers (RTNs), bank charters, and others are associated with the function of depository institution in the US
- Identity resolution of the counterparties to complex instruments for risk analysis is one of the most difficult tasks regulators have today

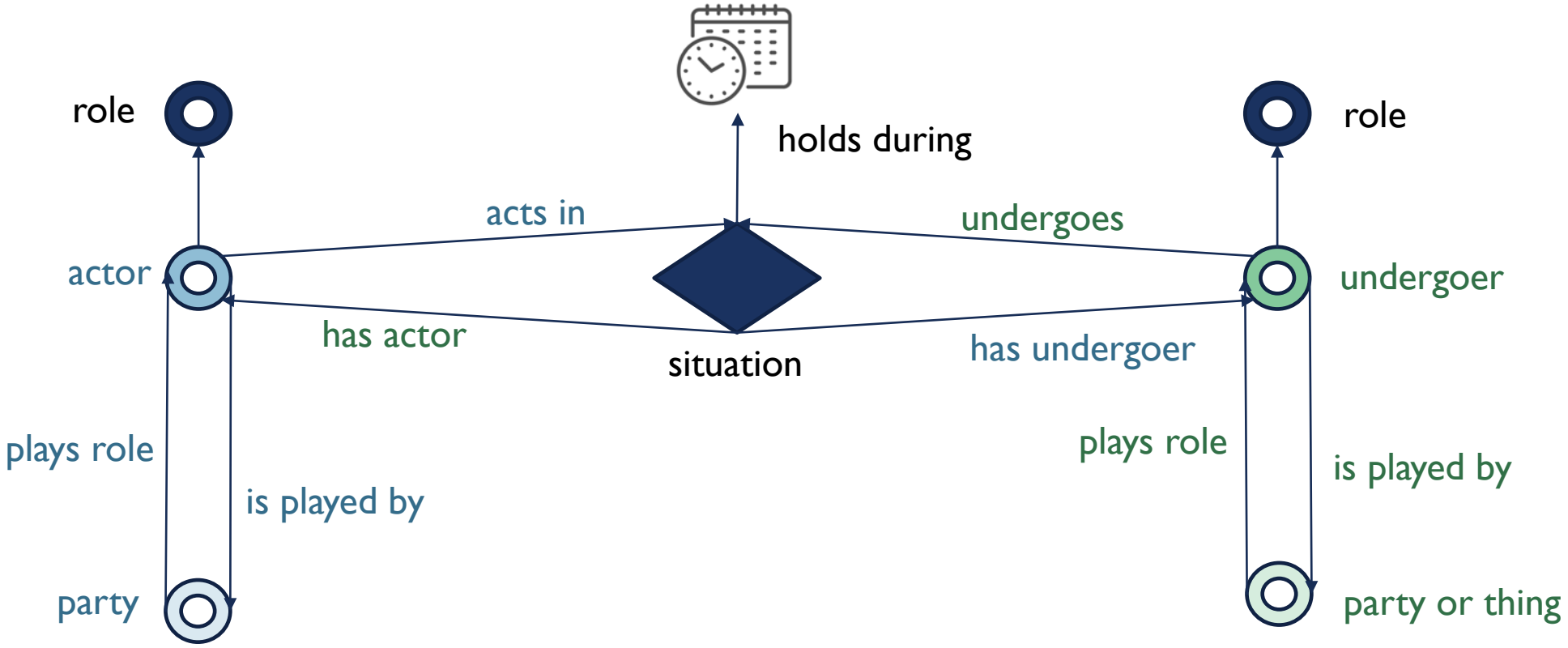
SOME RELATIONSHIPS ARE A BIT MORE COMPLICATED …

- A 'situation' is a setting, state of affairs/being, or relationship that is relatively stable for some period of time
- Examples include ownership, control, possession (which may or may not imply ownership), affiliation, beneficial ownership, board membership
- Situational analysis – enables traversal of these relationships, understand who owns who, who might know who, who might have influenced who, … who trades with who in more complex trading patterns, etc.
- Understanding these patterns combined with machine learning and other rule-based analyses allows us to
  - identify front running, insider trading, wash trading, other potential issues
  - roll up risk through ownership relations to provide the transparency needed to avoid the kinds of losses that Credit Suisse and others experienced due to Archegos Capital's recent implosion
  - explain results from complex learning and other analytics

HOW CAN WE BETTER REPRESENT OWNERSHIP?

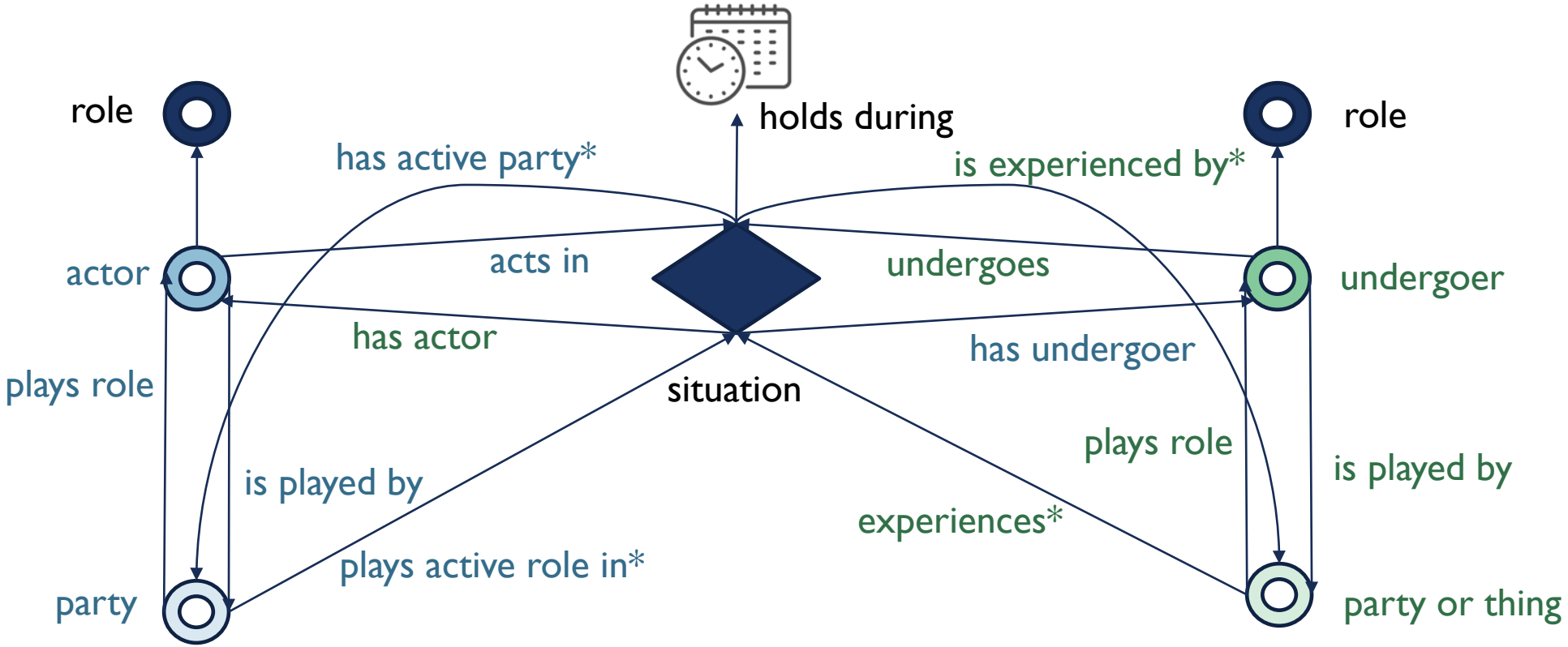


BASIC SITUATION



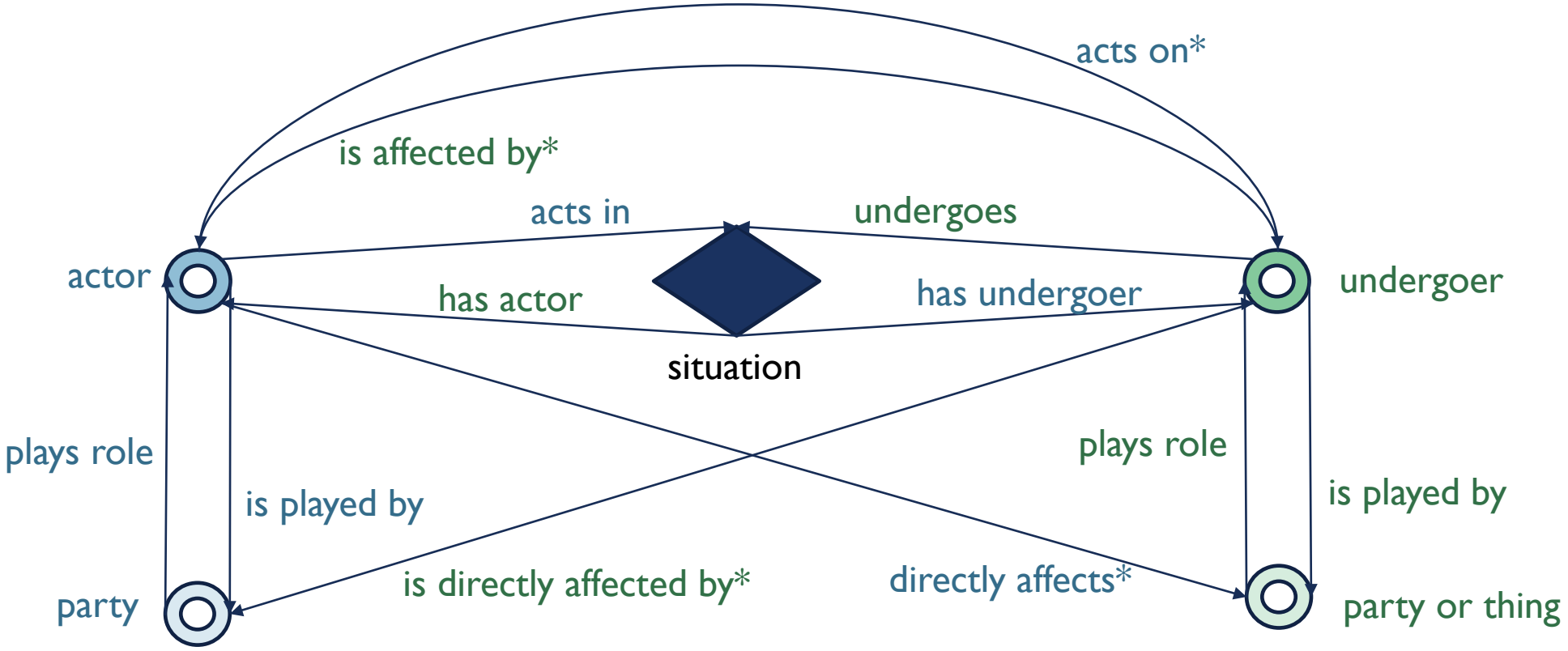
Pairs of properties are inverses of one another

EXTENDED SITUATION



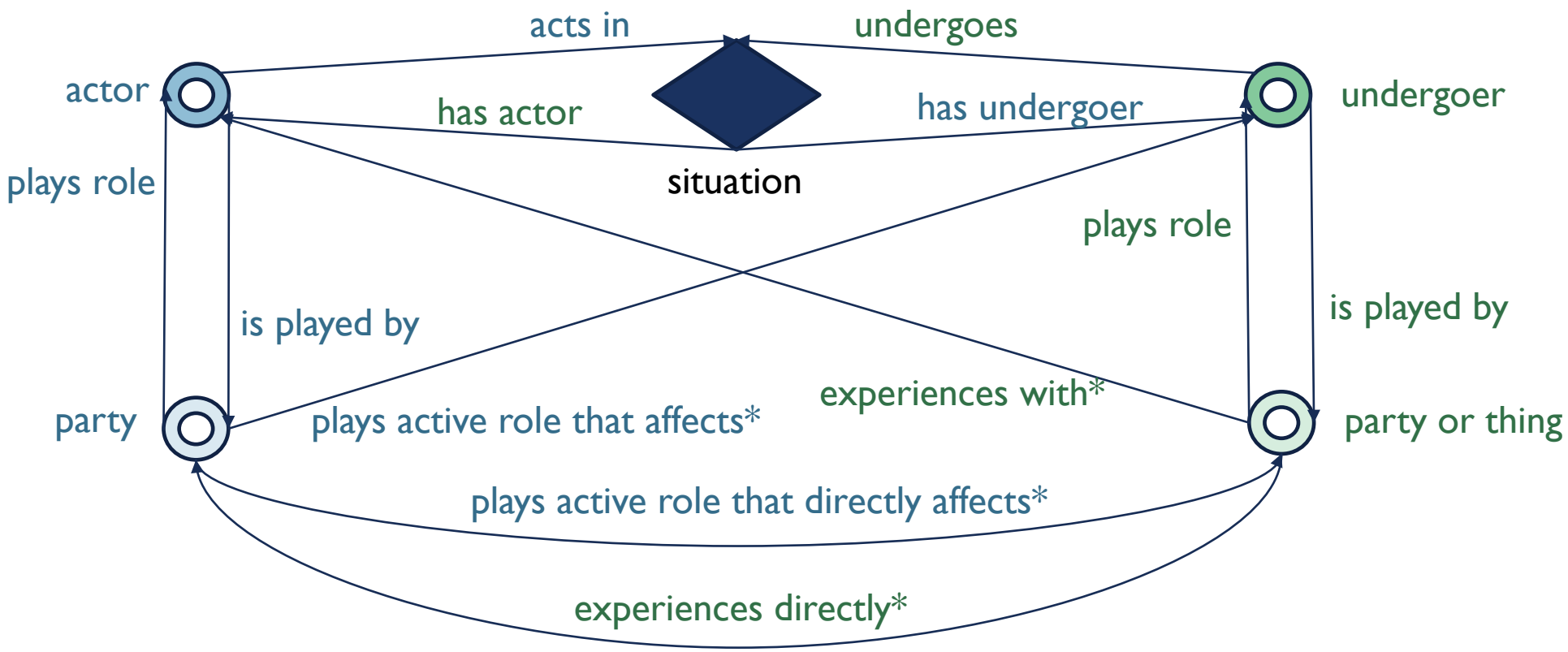
* Properties are chained; each pair of chained properties has a corresponding party to role inverse

SITUATION WITH ROLE TO ROLE / PARTY RELATIONS



* Properties are chained; Each pair of chained properties has a corresponding role to party inverse

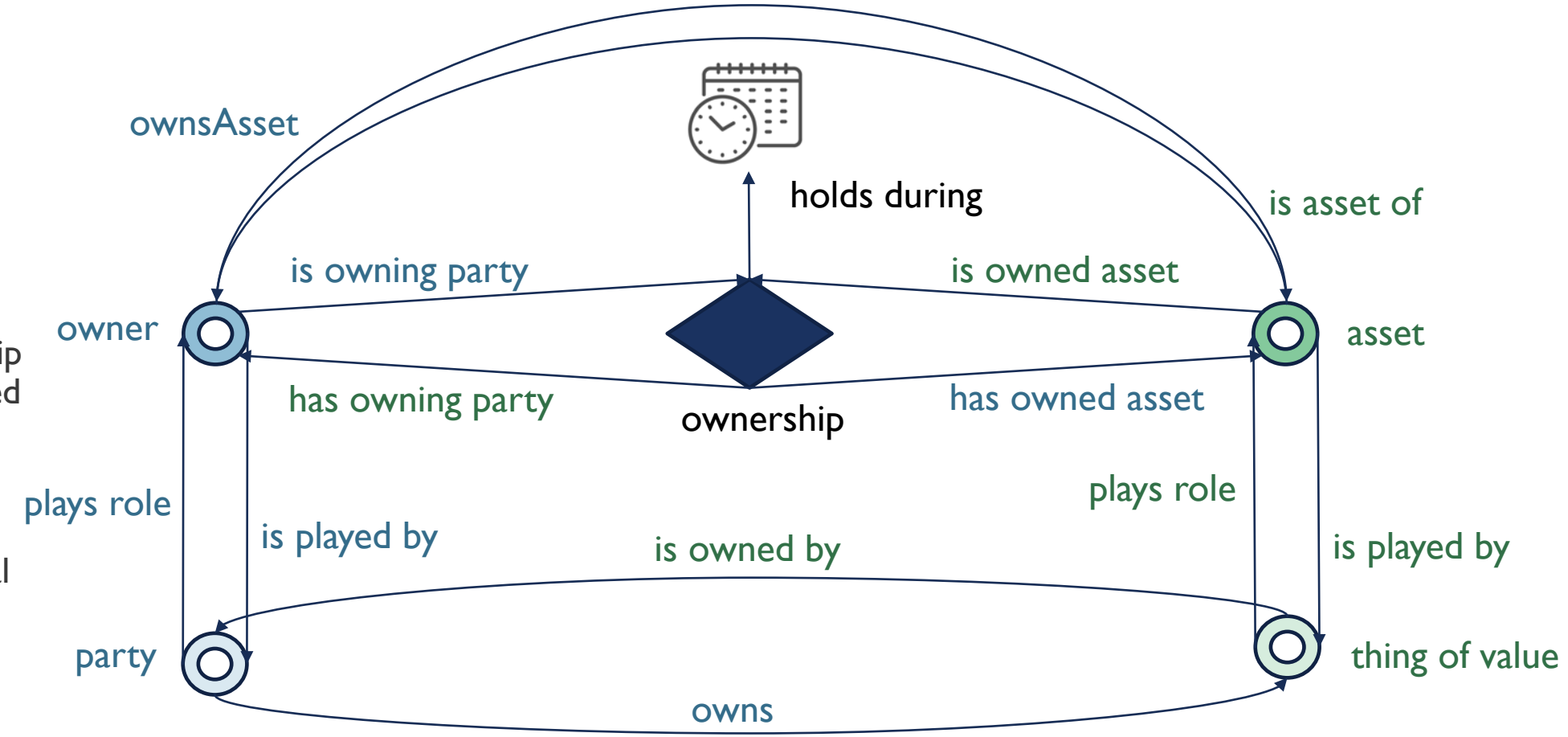
SITUATION WITH PARTY TO PARTY / ROLE RELATIONS



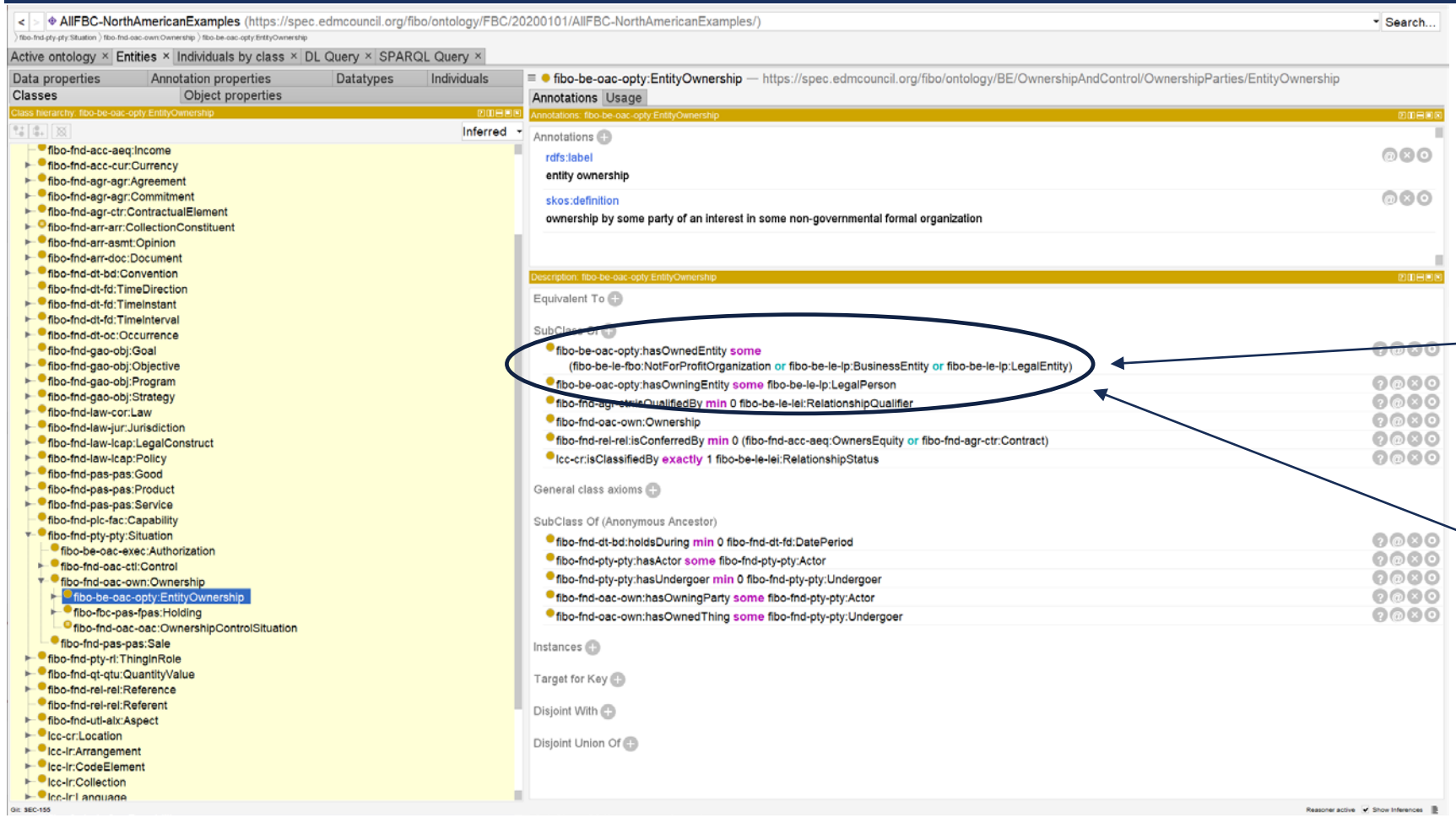
* Properties are chained; Each pair of chained properties has a corresponding party to role inverse

WHY MAKE THIS SO COMPLICATED?!?

- Ownership may be partial
- It requires evidence
- It may be beneficial through a broker or custodian
- Corporate ownership structures are layered thru subsidiaries
- Combined with identity resolution, this pattern is critical to understanding exposure across complex financial instruments



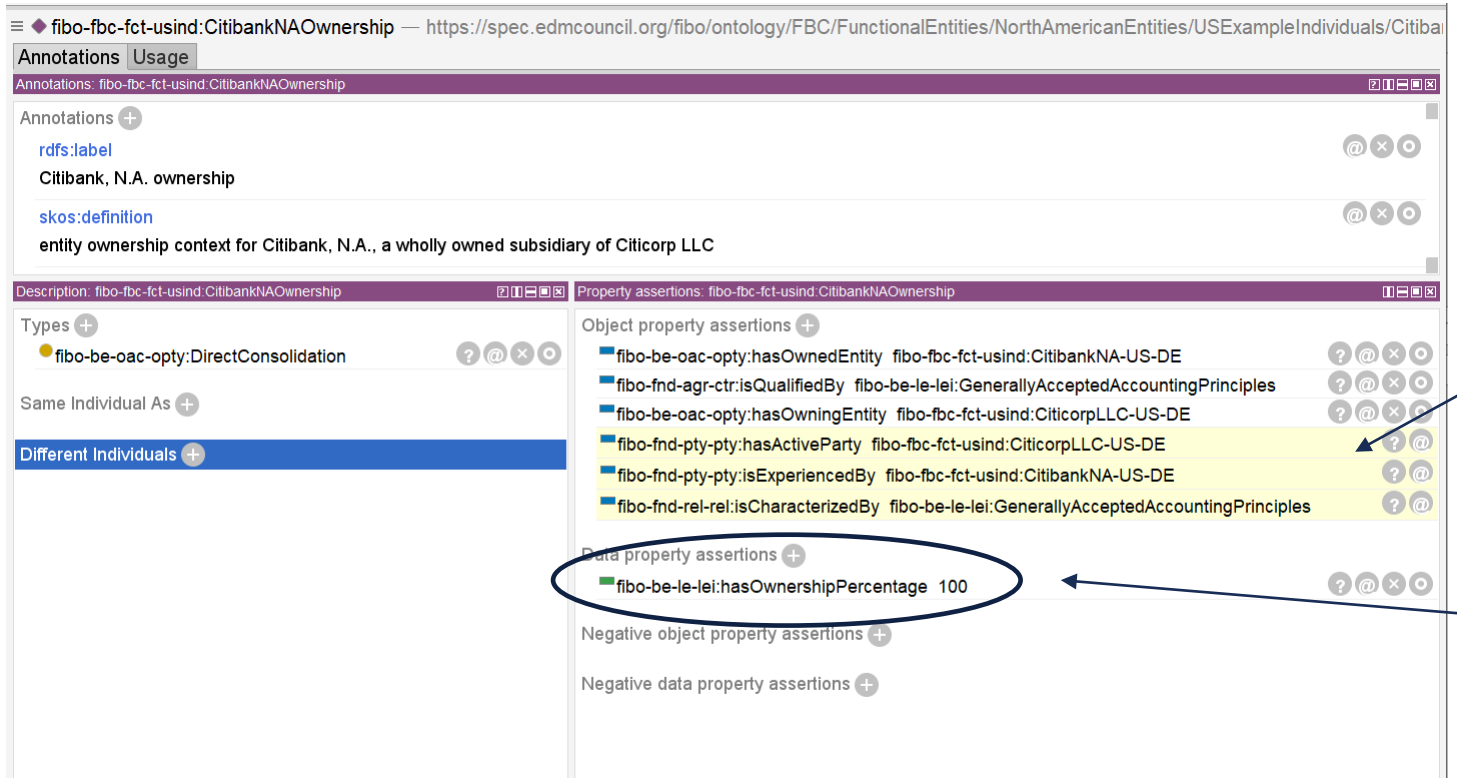
ENTITY OWNERSHIP



'has owned entity' is a subproperty of 'has active party', which is the chain 'has actor ○ is played by'

'has owning entity' is a subproperty of 'is experienced by', which is the chain 'has undergoer ○ is played by'

ENTITY OWNERSHIP EXAMPLE

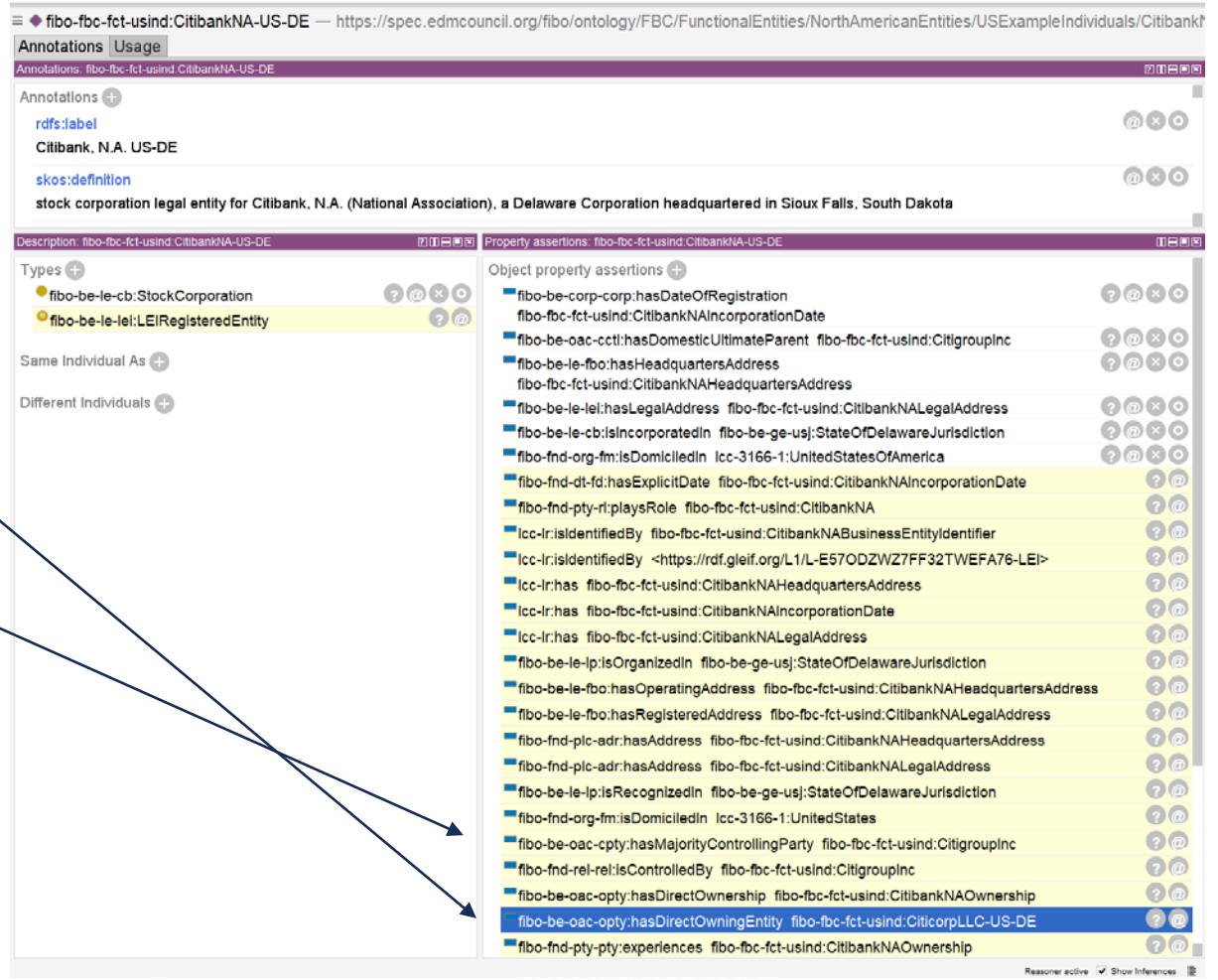


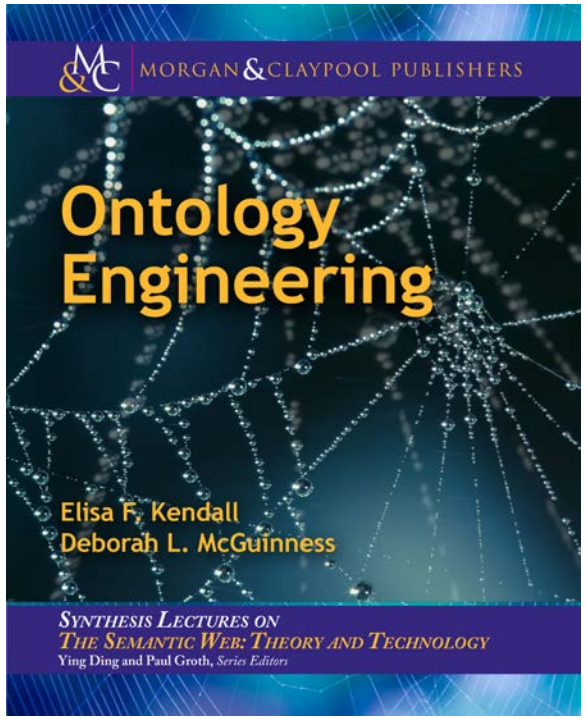
The inferences fill in the blanks

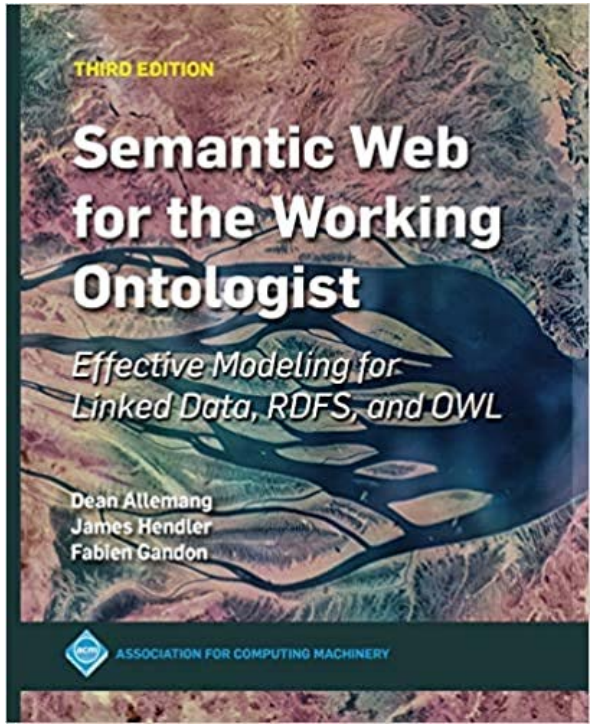
And we can add more details that are specific to the ownership situation, such as the percentage of the company owned

ENTITY OWNERSHIP EXAMPLE

- And, now, from the legal entity node for Citibank, N.A., now we can see that Citicorp LLC is its direct owner
- And through additional links in the graph, that it's majority owner is Citigroup, Inc.







THANK YOU

Elisa Kendall
Partner, Thematix Partners LLC
ekendall@thematix.com

Dean Allemang
CEO, Working Ontologist LLC
dallemang@workingontologist.com