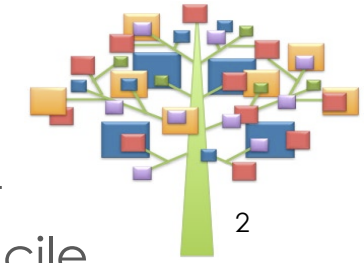
Commons Ontology Library – Digital Engineering Model Interoperability (DEMI)

onto:Nexus Workshop
30 January 2024

Elisa Kendall and Roger Burkhart
Thematix Partners LLC

About FIBO



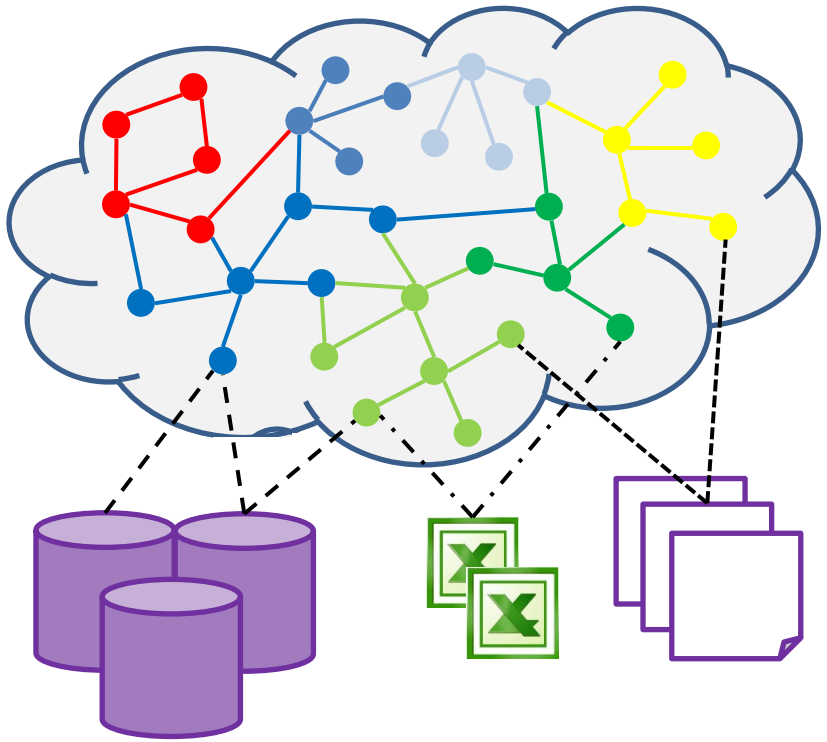
- The Financial Industry Business Ontology (FIBO) is an industry-level ontology that provides standard terminology, relationships, and logic designed to help reconcile disparate language defining financial instruments and related knowledge
- Initially developed by data management professionals from a 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

EDMCouncil

OMG OBJECT MANAGEMENT GROUP®

FIBO Financial Industry Business Ontology

Towards interoperability, decision support, new insights through semantics / knowledge graphs

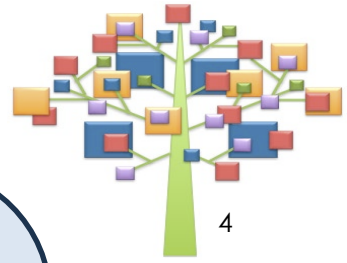


- **Data linkage and integration despite silos**
- **Open global reusable data standards**
- **Consistent, unambiguous definitions including the logic that distinguishes them for people and machines**
- **Highly expressive, flexible data schemas – alignment and interoperability based on meaning, with support for provenance, lineage, interdependencies by design**
- **Rich multi-level taxonomies, controlled vocabularies**
- **Pattern-based architecture to enable AI / machine-learning with explanations**
- **Much higher accuracy and provenance for question answering with large and custom language models (LLMs)***

** https://arxiv.org/pdf/2311.07509.pdf*

Financial Industry Business Ontology

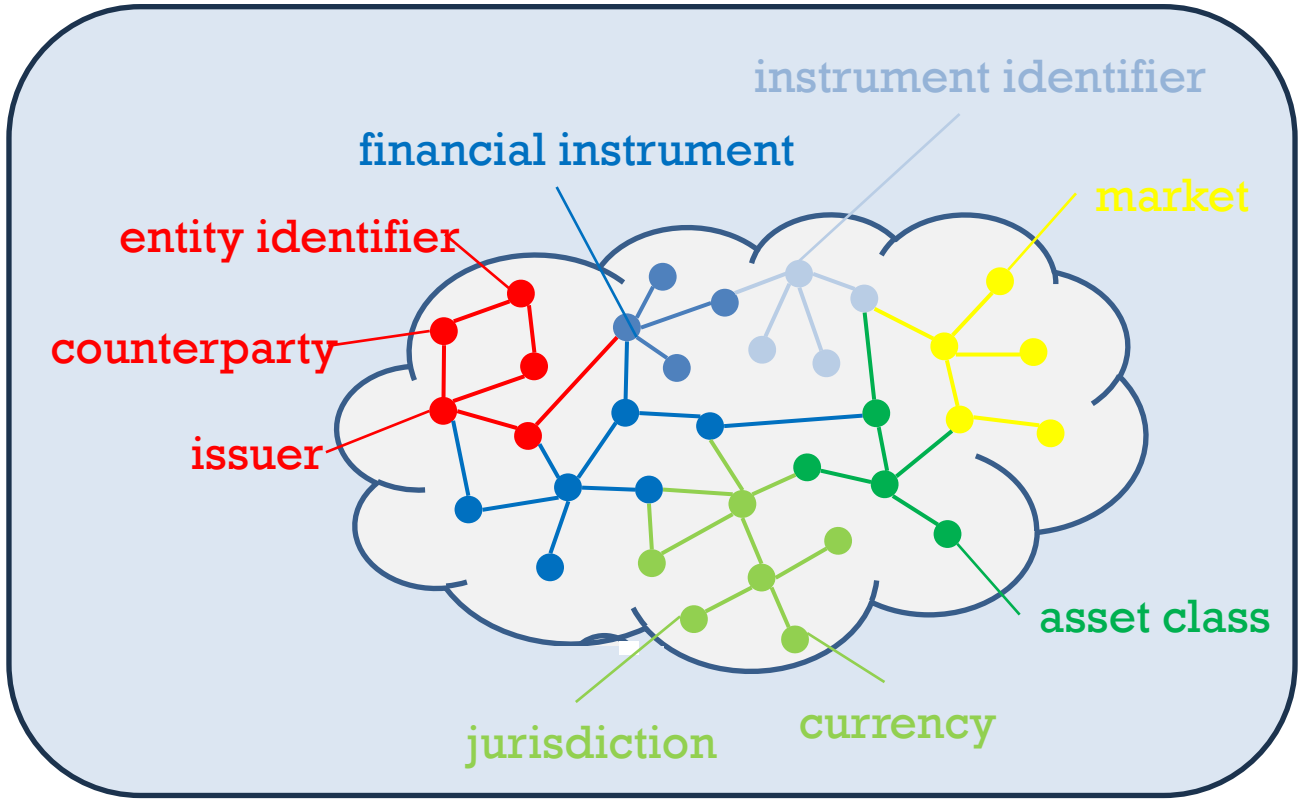
FIBO
Financial Industry Business Ontology



GLEIF

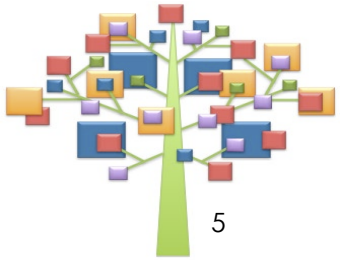
OMG Standards Development Organization

ACTUS
ALGORITHMIC CONTRACT TYPES UNIFIED STANDARDS



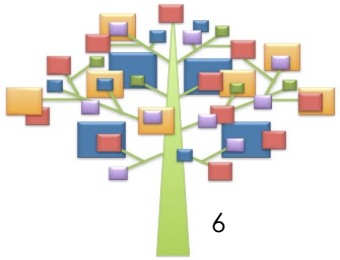
- Consistent framework for defining entities, the roles they play, relationships between them, with deep instrument hierarchies, securities master data terms, accounts …
- Semantically rich basis for extension, enabling efficiency, transparency
- Forward-looking, pattern driven design supports artificial intelligence / machine learning / LLMs

Motivation for the Commons Ontology Library



- Patterns increase the quality of any large ontology, increase efficiency and enable better understanding for new users
- Modeling patterns have been used in the Financial Industry Business Ontology (FIBO) for several years, longer in some cases
- Work with millions of records of data from member institutions has proven that patterns improve efficiency in mapping to source data and in query development for knowledge graph analysis
- Insights gained have led to development of new, improved patterns over the last several years
  - Recent updates to FIBO focus on increasing pattern usage, limiting ad hoc approaches where possible
  - Many of the patterns are not unique to the finance industry

Multiple Vocabulary Facility (MVF)

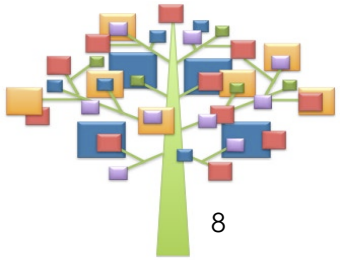


- OMG identified challenges in enabling modelers, particularly those developing business process models, business architectures, model-based systems engineering, and ontologies to work in their native natural language or nomenclatures
- MVF is designed to fill the gap – to enable people using MOF/UML for modeling to apply vocabularies that work for the modeler and domain, and allow others to reuse the same model, but in their own language to facilitate understanding
- Key requirements include
  - a standard means for representing the relationship between model elements and the corresponding business concept definitions and multiple term sets
  - a standard means for accessing terms of a selected vocabulary for a model element and interpret a selected vocabulary term as a reference to a particular model element representing that concept
  - a standard form for import and export of vocabularies
  - and the ability to use this functionality not only for user models but for any MOF-based metamodel or profile, such as UML, BPMN, SysML, and others

Need for a common library for MVF

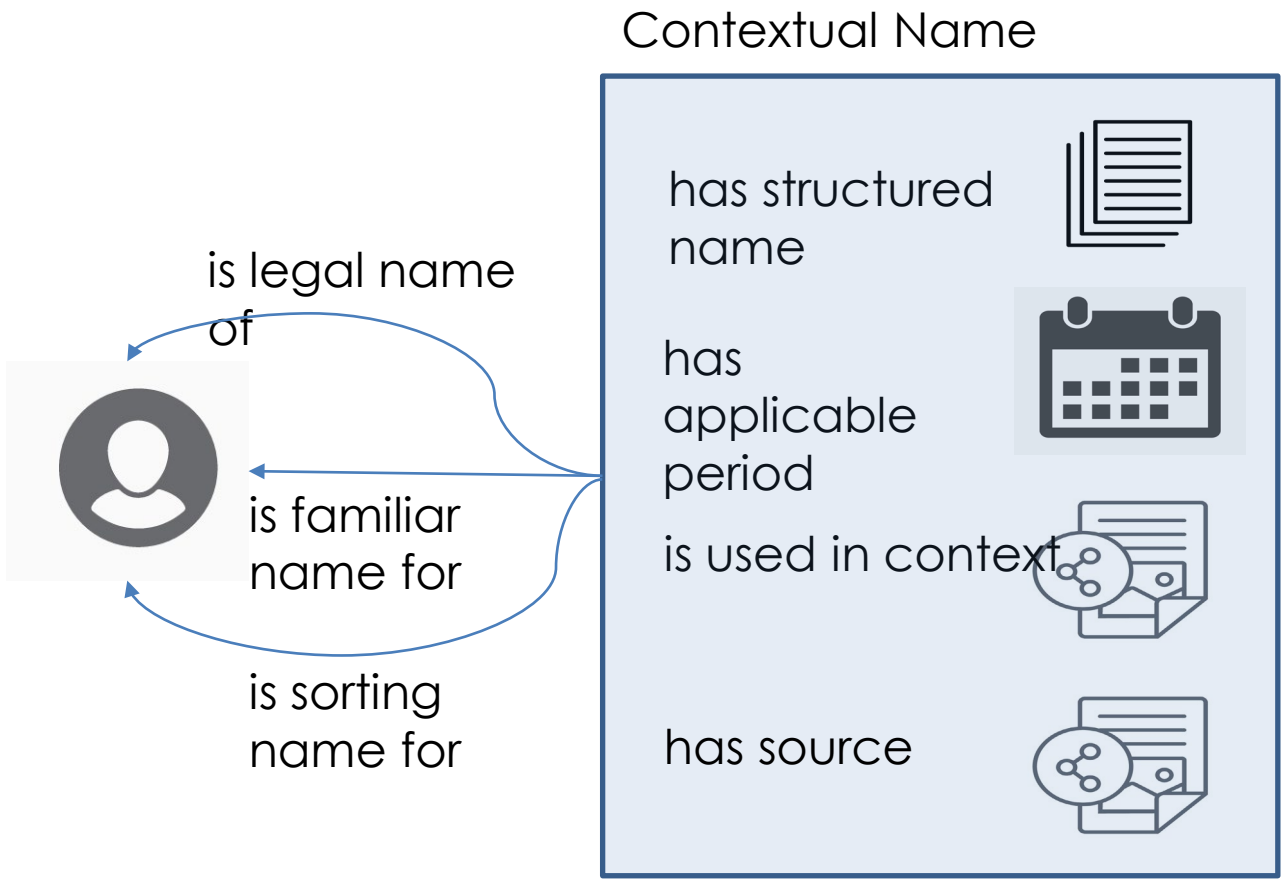
- MVF includes ontologies for vocabulary representation based on ISO 1087, ISO 11179 and related standards
- Community members with cross-domain expertise found that
  - Patterns needed for MVF were also needed across several OMG efforts (*e.g.*, FIBO, Languages, Countries and Codes (LCC), an emerging retail industry ontology (RIO), a joint effort with IEEE on an ontology for service robots (RoSO), and others)
  - Most were not found in existing ontology libraries as standalone ontologies
  - An initial Commons Ontology Library 1.0 Specification was published in August 2023
  - Four additional ontologies were added in the Commons 1.1 revision, currently in the final approval/publication process
  - The library is designed for extension - ontologies can be incorporated through contributions from existing, well-tested work with mapping to other well-known ontologies where possible

“Simple” patterns included in the library



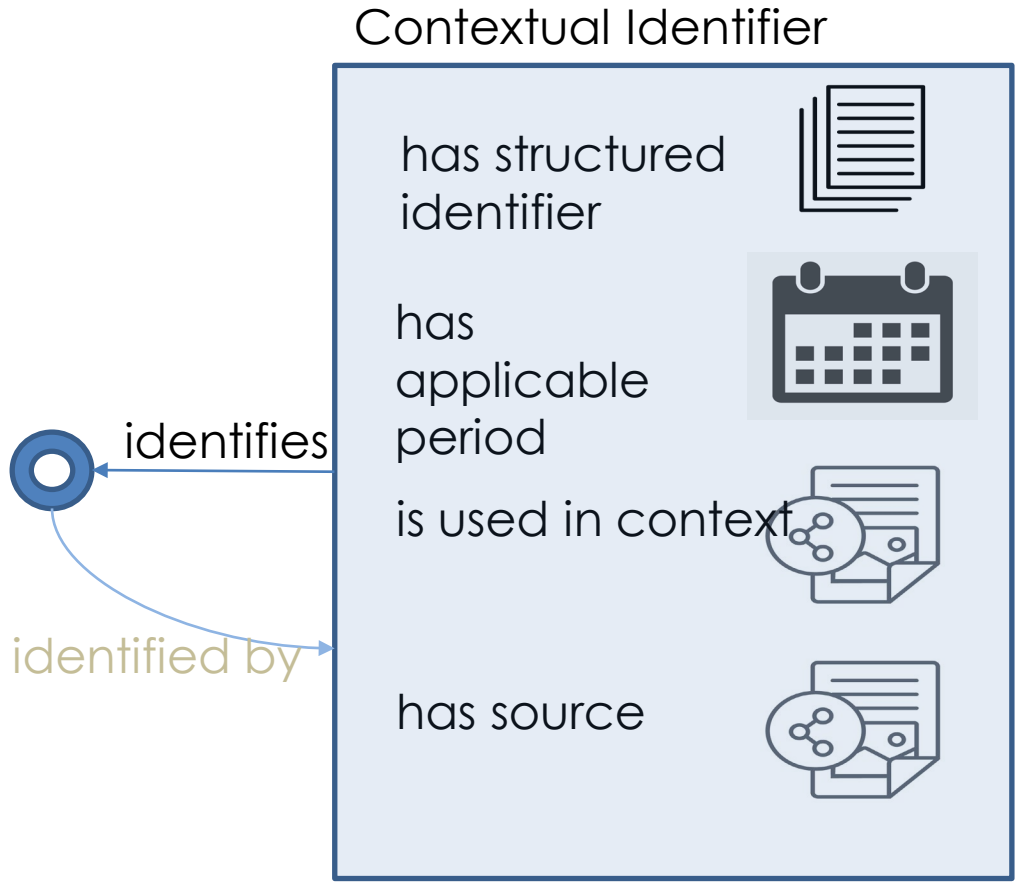
- Assigning names to things – to facilitate name resolution across large corporations – for any 360° view of your customer in a large enterprise, regulations in banking known as ‘know your customer’ (KYC), etc.
- Identifying things – identity resolution in finance is typically more challenging than name resolution, but is important for many industries – retail, healthcare / pharma, manufacturing …
- Relating codes and code sets to things – codes assigned to industries, markets, places, products, 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

Designations / Naming



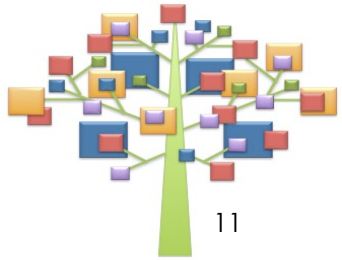
- Commons includes simple and contextual designation support
- Patterns apply to people, organizations, places, other things for which designations are needed
- Key use cases include normalizing customer / supplier names, names of medicinal products and substances
- Facilitates mapping to various resources for diverse applications
- Supports normalization across resources, contexts and reusability in general

Identification



- Commons includes simple, contextual, and structured identifier support
- Structured identifiers – including identifiers that are composites, *e.g.*, ISIN (with country code) vs. NSIN (without country code)
- Key use cases include legal entity identifiers for counterparties, securities identifiers, identifiers for products in agriculture (seed to table), identifiers for medicinal products and the substances they contain
- Features of these kinds of identifiers:
  - Registered with some registration authority
  - Minted according to some sort of scheme

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 2 M+ records is available on data.world using the GLEIF ontologies

LEI Reference Data

Back to search results

Citibank, National Association

Current Data | Change History

legal entity identifier

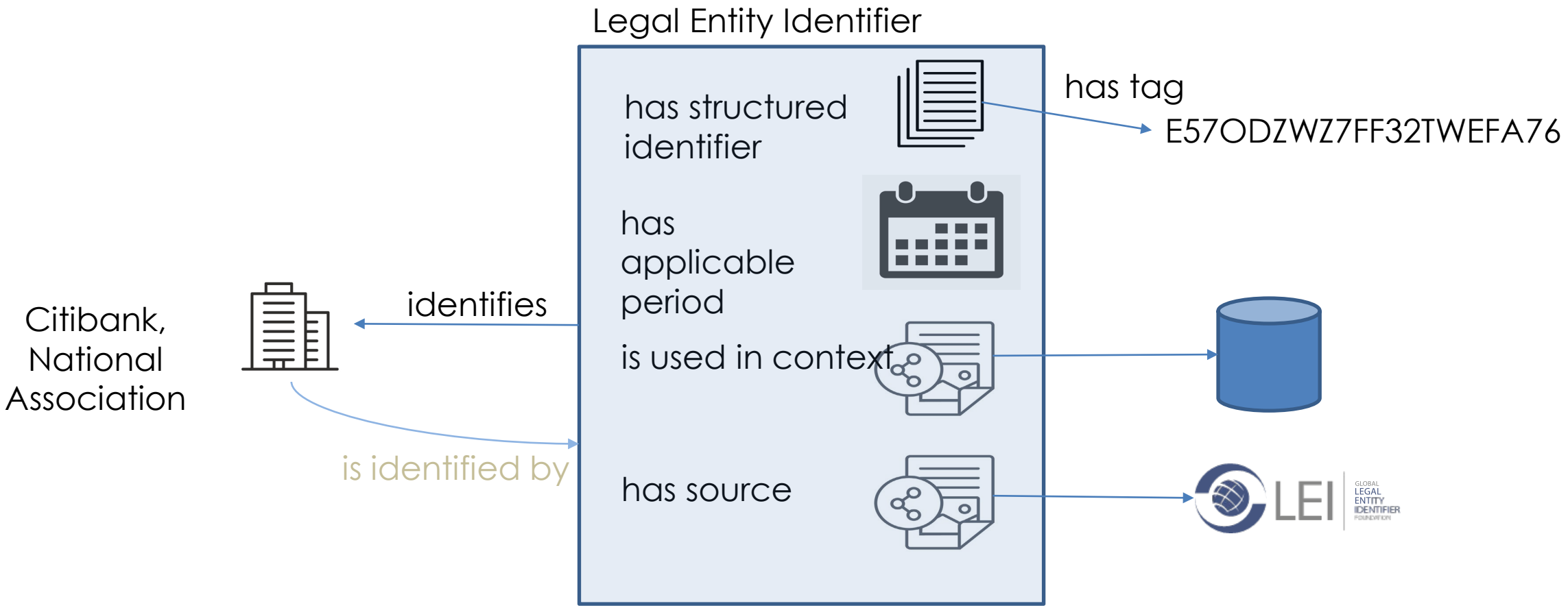
validation source

LEI Code E57ODZWZ7FF32TWEFA76

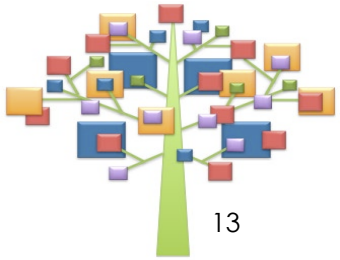
Hide

(Primary) Legal Name	Citibank, National Association
Registered At	Registry of FDIC-insured banking institutions (Federal Deposit Insurance Corporation) Registry of FDIC-insured banking institutions (Federal Deposit Insurance Corporation) United States of America RA000744
Registered As	7213
Jurisdiction Of Formation	US-SD
Entity Legal Form	Temporary Code (8888) No alternative legal form description provided
Entity Status	ACTIVE

Identifier Example

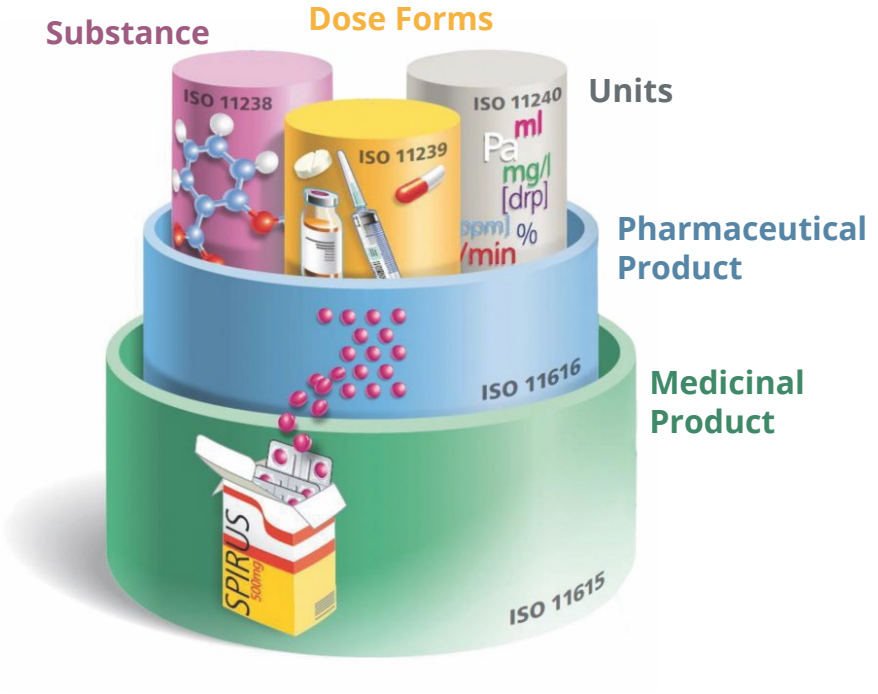


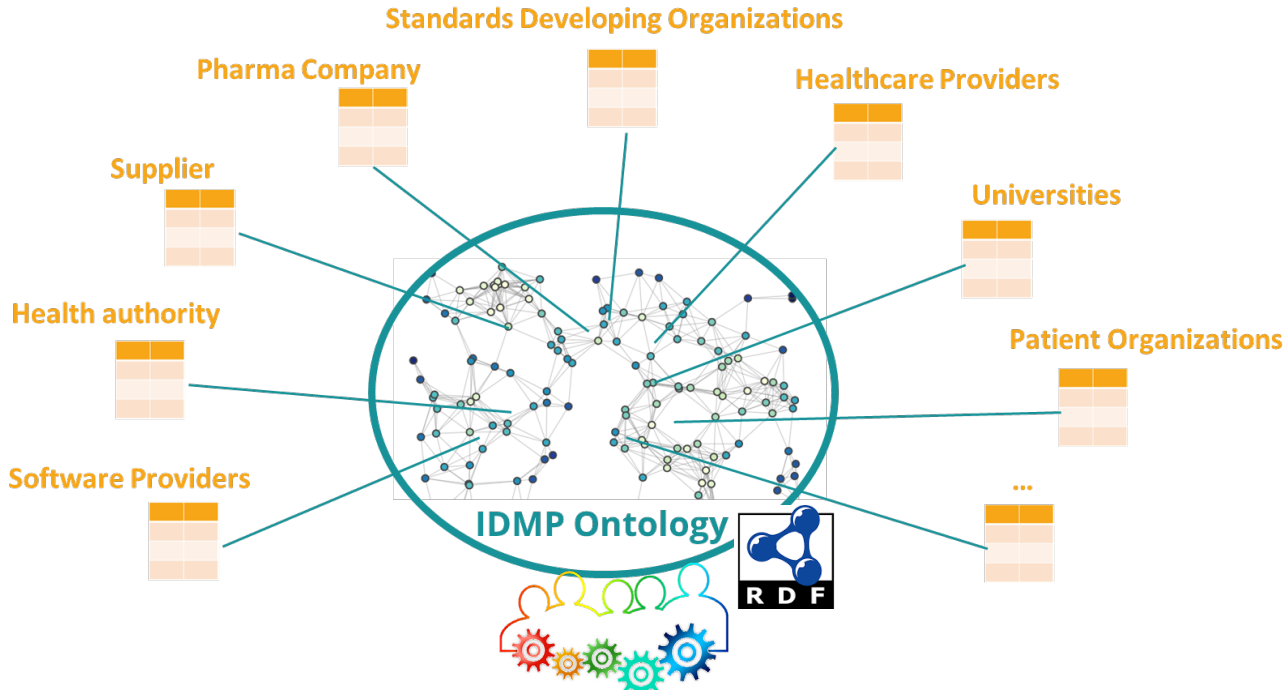
Other “simple” patterns in commons



- A basic annotation library that declares commonly used annotation properties
- Collections
- Basic dates and date periods with a mapping to W3C Time in OWL 2
- A text datatype that allows support for a combination of plain and language-tagged strings

Identification of Medicinal Products Ontology (IDMP-O)



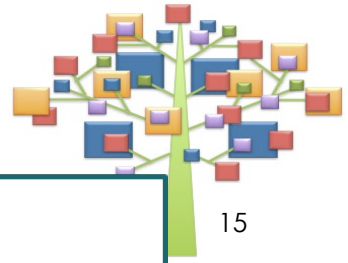


ISO IDMP Standards

Collaborative implementation creates interoperability by design

The IDMP Ontology provides a universal implementation of the IDMP product data model as a common language to effectively bridge the gap between people, processes, and systems.

IDMP Project Structure



IDMP Ontology Executive Advisory Board

Strategic guidance on long term roadmap (1 every 2 months)

Senior executives from founding pharma companies and key industry or regulatory stakeholders.

Project Steering Team

Provide direction and make strategic decisions on quarterly project priority tasks (at least quarterly)

Sponsoring organizations

BAYER · NOVARTIS · Roche · MERCK · GSK · Boehringer Ingelheim · Johnson & Johnson · AMGEN · AstraZeneca · abbvie · Pfizer · ...

Project Core Team

Collaborate and deliver on project scope and target results (bi-weekly)

People assigned from the Steering Team + WS team leads + selected key experts

STREAM	STREAM	STREAM	STREAM
Ontology Governance & Hosting	**Ontology Development**	**Use Case Implementation**	**Value Capture & Use Case Specification**
EDMCouncil	EDMCouncil · OSTHUS	ACCURIDS · Chemantics	OSTHUS

Project Community of Interest

Present project results and insights for a discussion in an open forum of innovators (every 6 weeks)

whole team with an open list of interested people from any organization

Requirements | Implementation Feedback

Project Lead & Management

Gerhard Noelken, Sheila Elz
Heiner Oberkampf, Jan Kroh,
Ivana Miljkovic

Pistoia Alliance

- Agile Project Organization
- Marketing Communication
- Stakeholder Management

Interested Parties

Regulatory Agencies

FDA · G-SRS · EU-SRS · European Medicines Agency · SPOR · Uppsala Monitoring Centre · CBG MEB · College ter Beoordeling van Geneesmiddelen · World Health Organization

Standards Development Organizations

ISO · HL7 · SNOMED CT

Government Initiatives

PH Trade Associations DIA

Projects UNICOM

Not-for-profits CTADHL

Vendors
E.g., Veeva

Accumulus synergy

Use Cases and Approach

implementation

Substance Identification and Roles
Active moiety, ingredient strength & chemical groupings

Regulatory and Manufacturing
Enabling interoperability between manufacturing (bottom-up) and IDMP/labelling (top-down) perspective

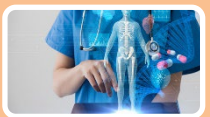
Therapeutic Indication
Linking medication to clinical particulars

Jurisdiction-agnostic Medicinal Products
Global Medicinal product that industry can refer to without any regulatory-specific data

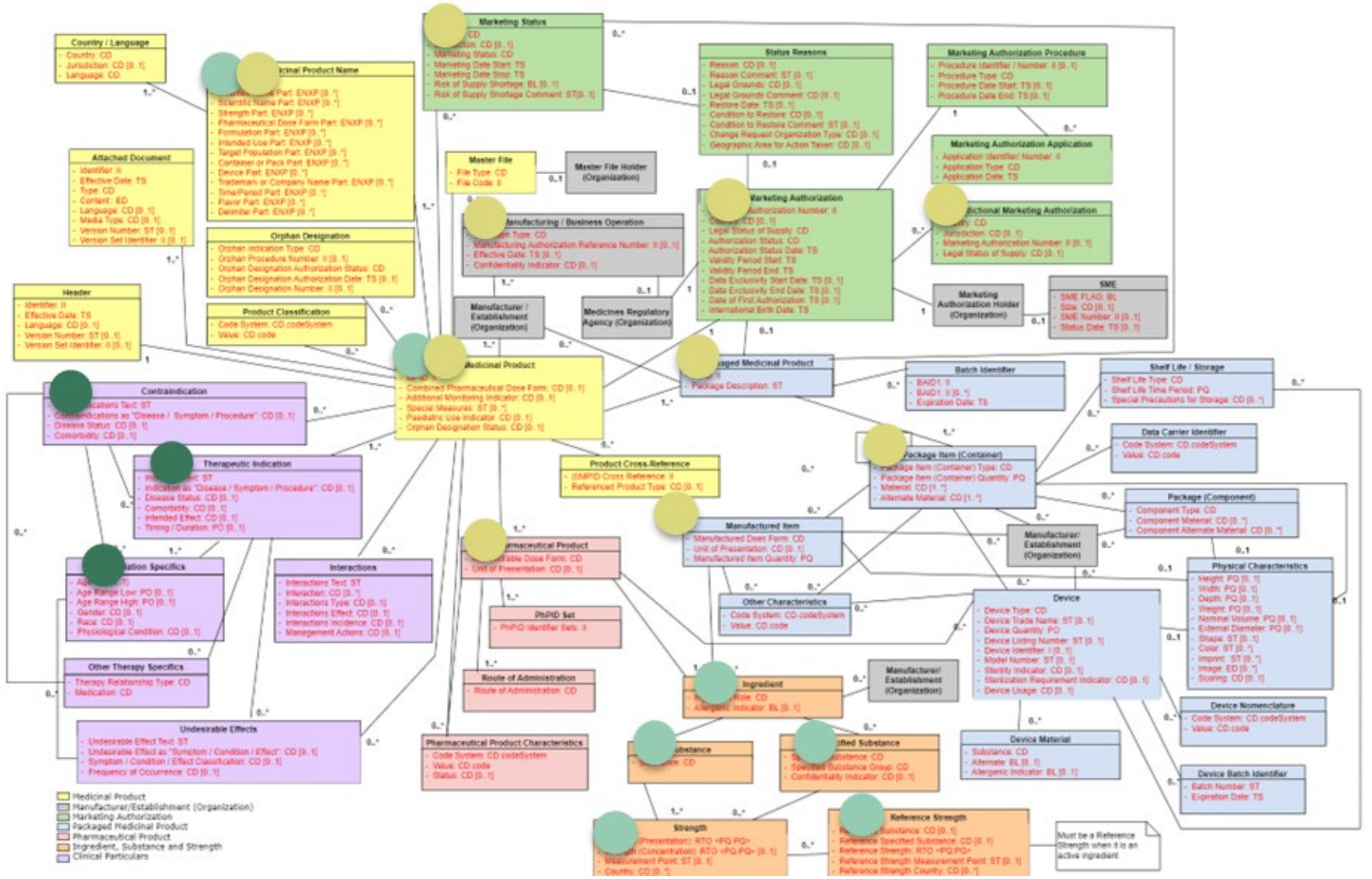
Falsified Medicines Directive
Integrated data for mandatory EMA reporting

backlog

Pharmacovigilance
Global Impact Assessment of Safety Risks Across the Product Life-Cycle

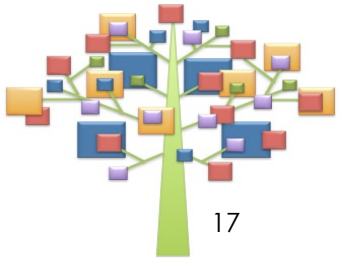


Clinical and Regulatory
Enabling interoperability between ClinOps and Regulatory incl. reference to CDISC



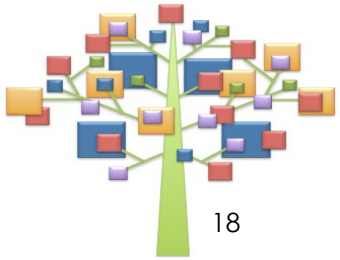
For more information, see https://www.pistoiaalliance.org/projects/current-projects/idmp-ontology/

Observations from reusing Commons & MVF for IDMP-O



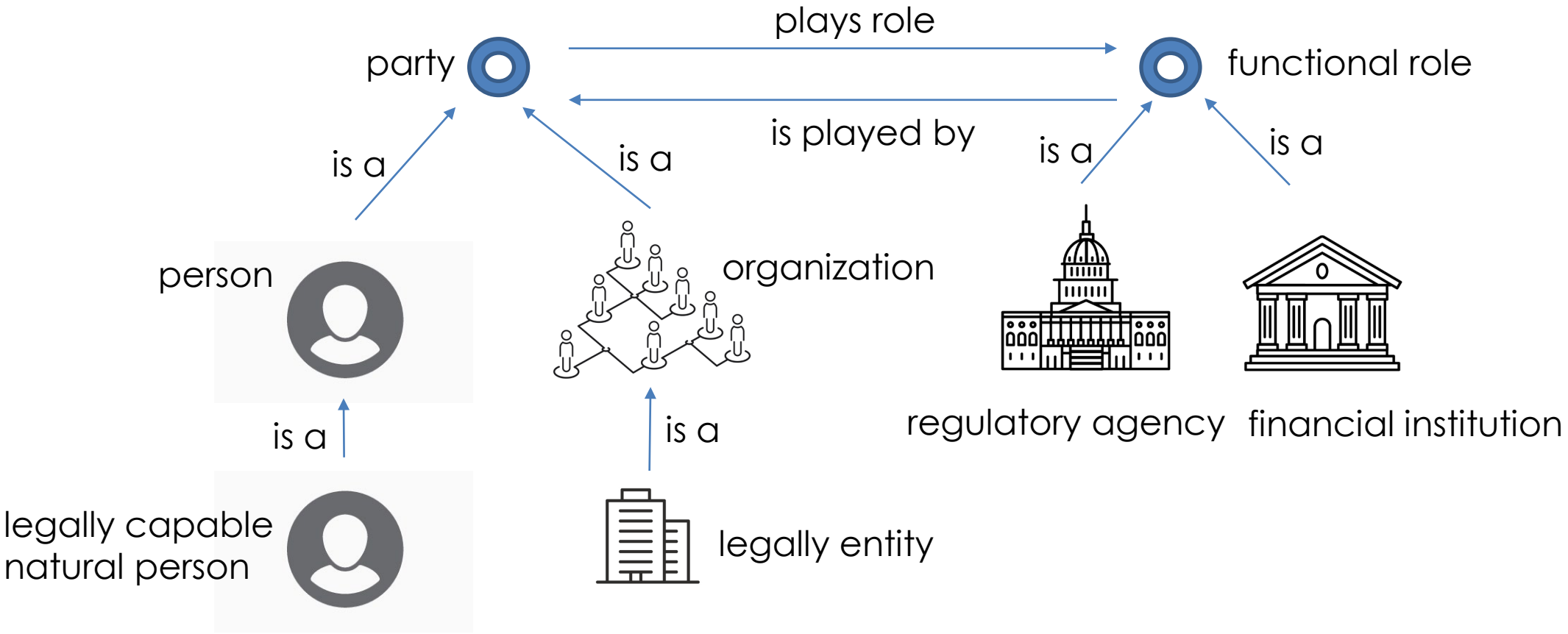
- The IDMP-O ontology consists of over 800 classes, over 400 relationships and over 100 attributes
  - More than 50% of the classes are from Commons and MVF, and over 65% of the relationships and attributes, or from planned additions to Commons
  - Reuse saved more than a year in the development process
- Pharma team users recognize the patterns and have mapped them to internal data in addition to the examples and public data mappings created
- Pharma regulators are considering reuse of the controlled vocabulary pattern internally
- Use of these patterns provides the building blocks for machine learning, natural language processing, custom and large language models

"More complicated" patterns included in the Commons 1.1 specification

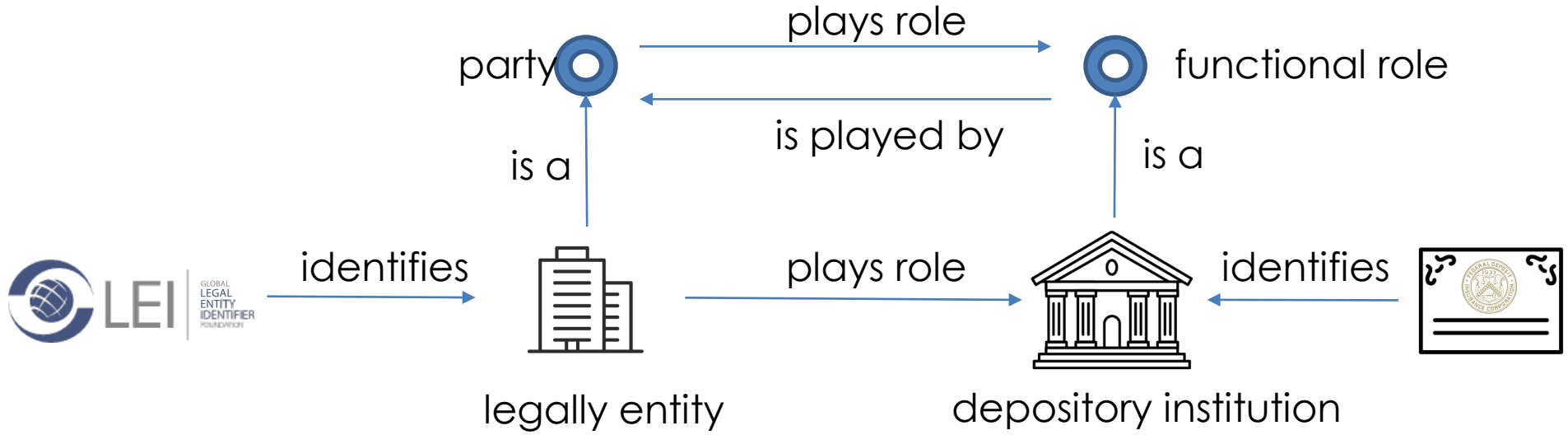


- Quantities and units – mappable to QUDT, UO, the NASA Sweet ontologies, SysML v2 (developed with support of SysML v2 and NIST)
- Associating parties (people and organizations) with the roles they play
- Situational analysis / reified relationships – linking parties and roles to situational patterns that are time bound (e.g., ownership, control, authorization, membership, employment …) and for pharma – connecting complex substances to the roles they play in medicinal products
- All of these have been tested in FIBO and in an external project for the Identification of Medicinal Products (IDMP), and others

Parties vs. Roles

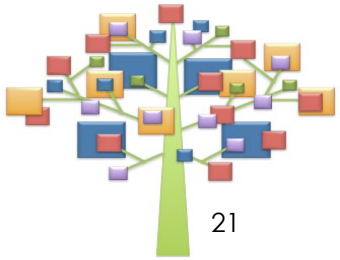


Identifiers for Parties or Roles?



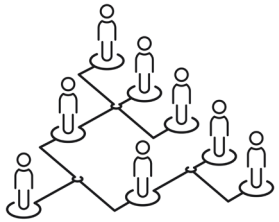
- LEIs, business identifiers issued by state governments, and others identify legal entities directly
- FDIC certificates, Routing Transit Numbers (RTNs), bank charters, and others are associated with the function of depository institution in the US
- Identity resolution of the counterparties to complex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

Some relationships are even more complicated

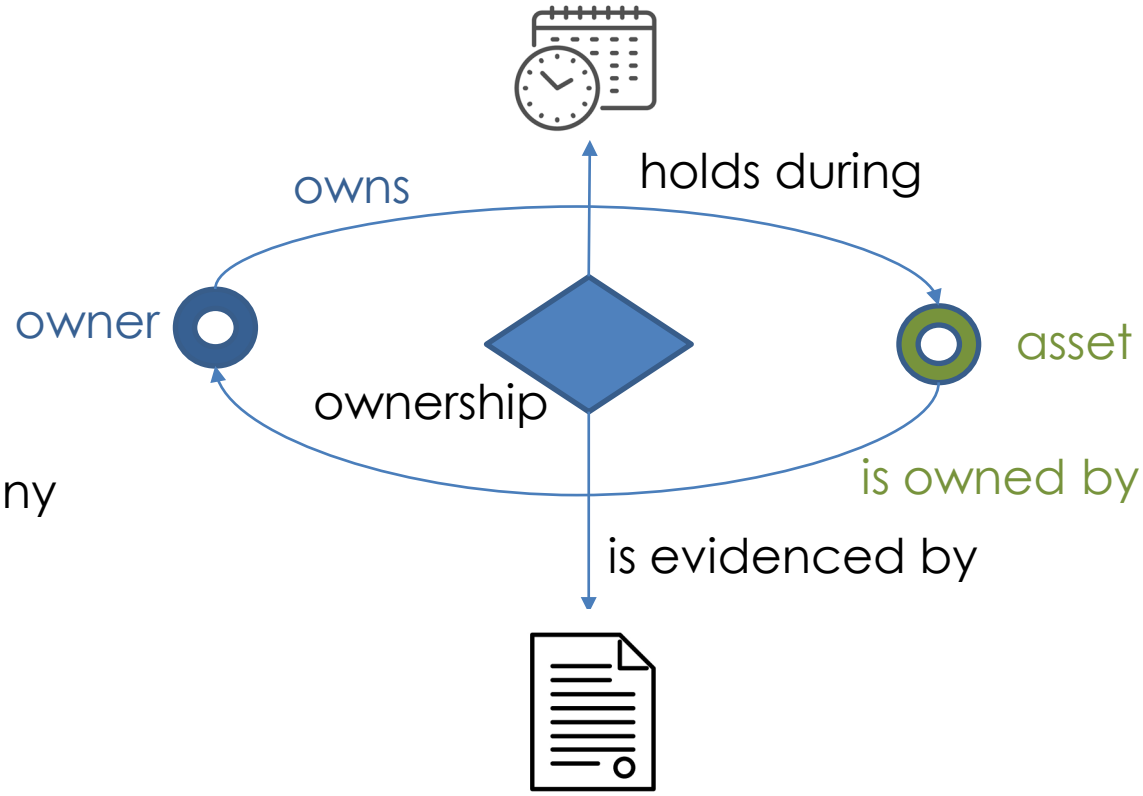


- A ‘situation’ is a setting, state of affairs/being, or relationship that is relatively stable (holds) for some period of time
- Examples include ownership, control, possession (which may or may not imply ownership), affiliation, beneficial ownership, board membership, employment, …
- Situational analysis – 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 implosion earlier this year
  - explain results from complex learning and other analytics

How can we better understand ownership?

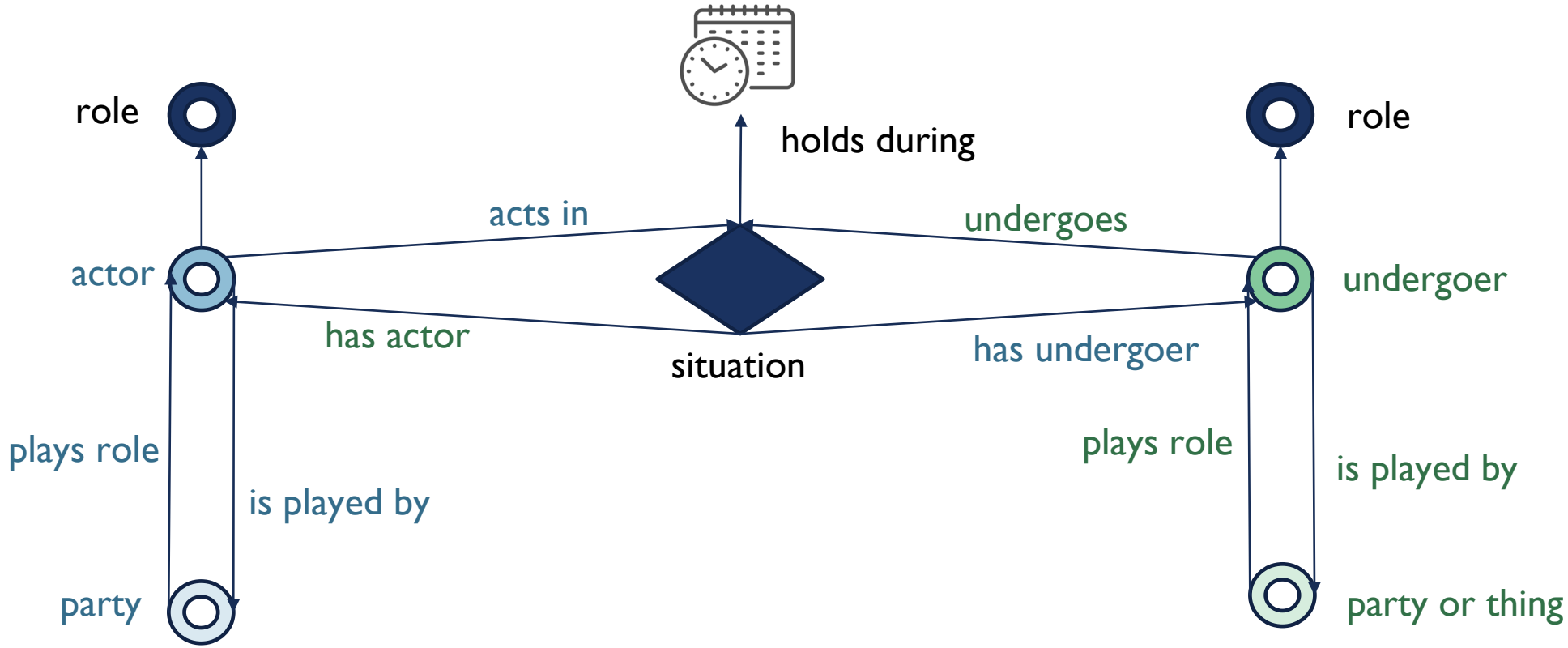


holding company



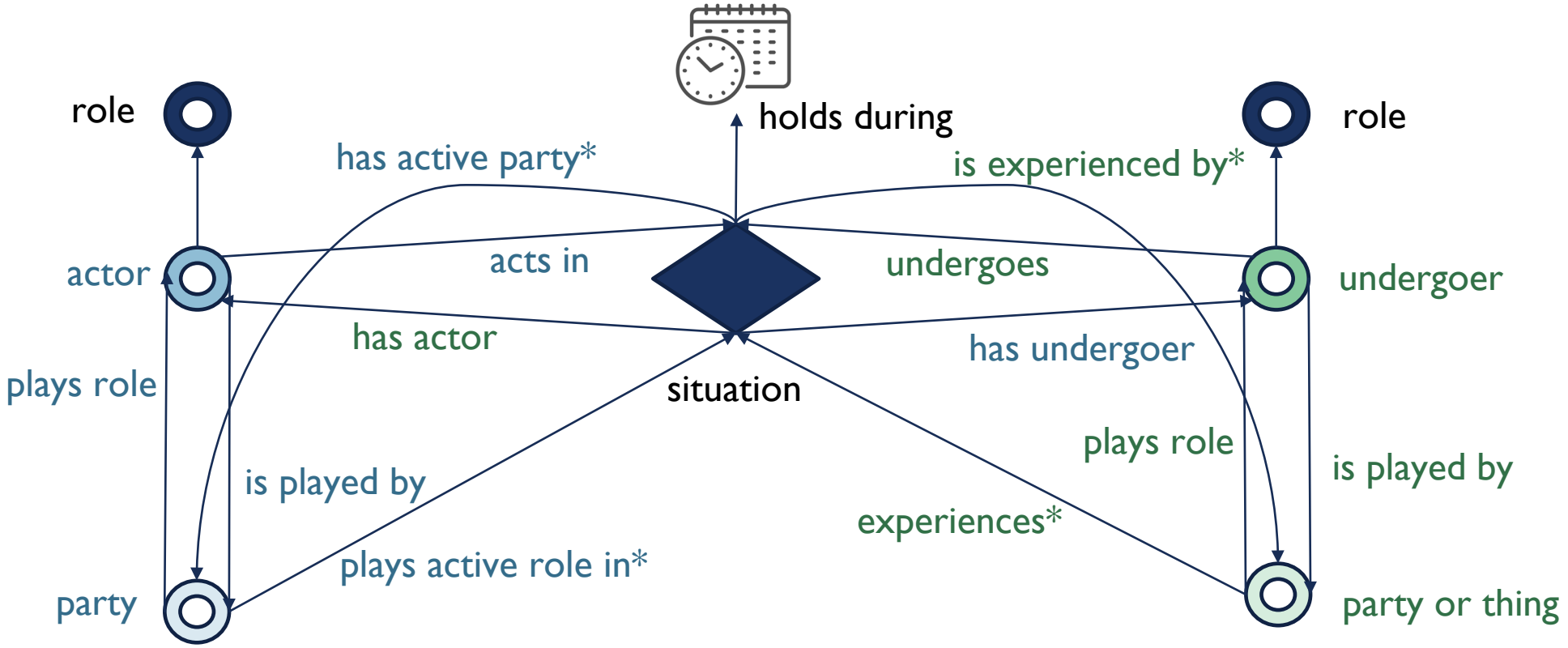
financial institution

Basic situation



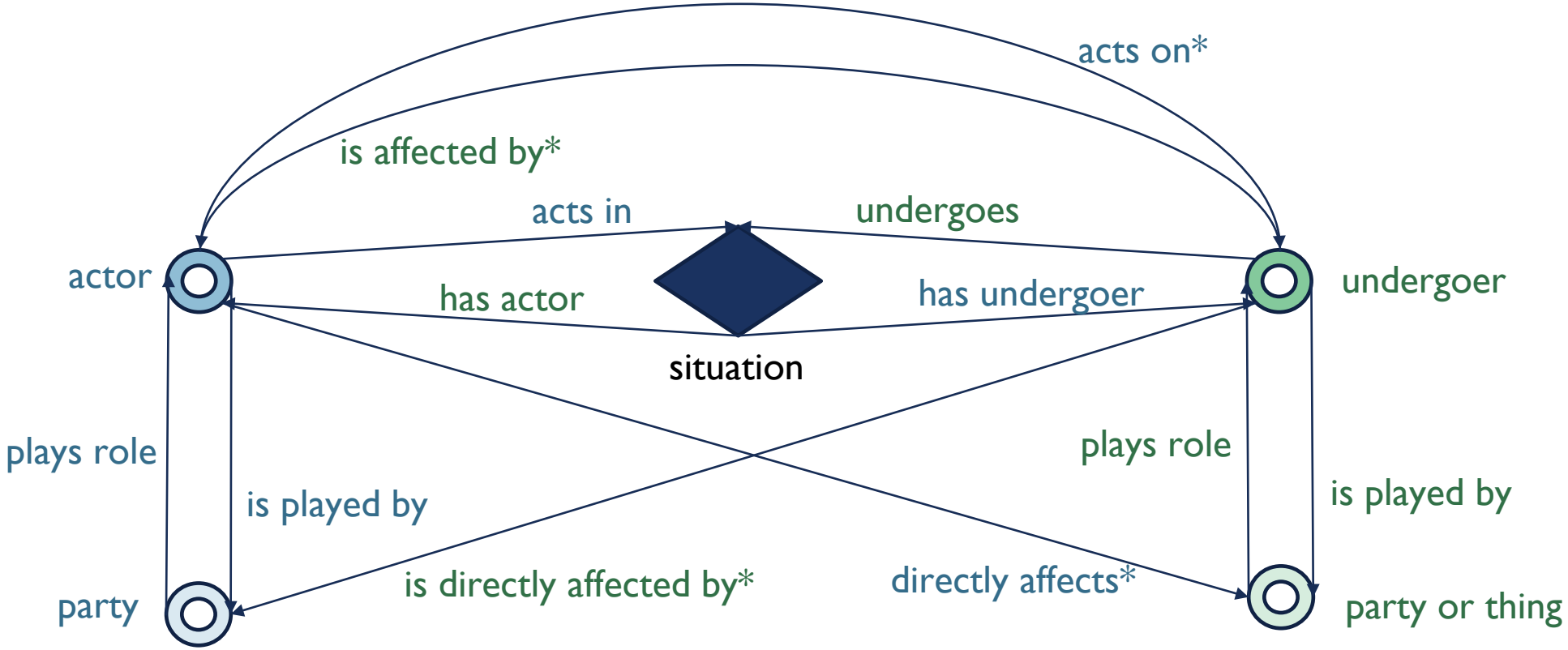
Pairs of properties are inverses of one another

Extended situation



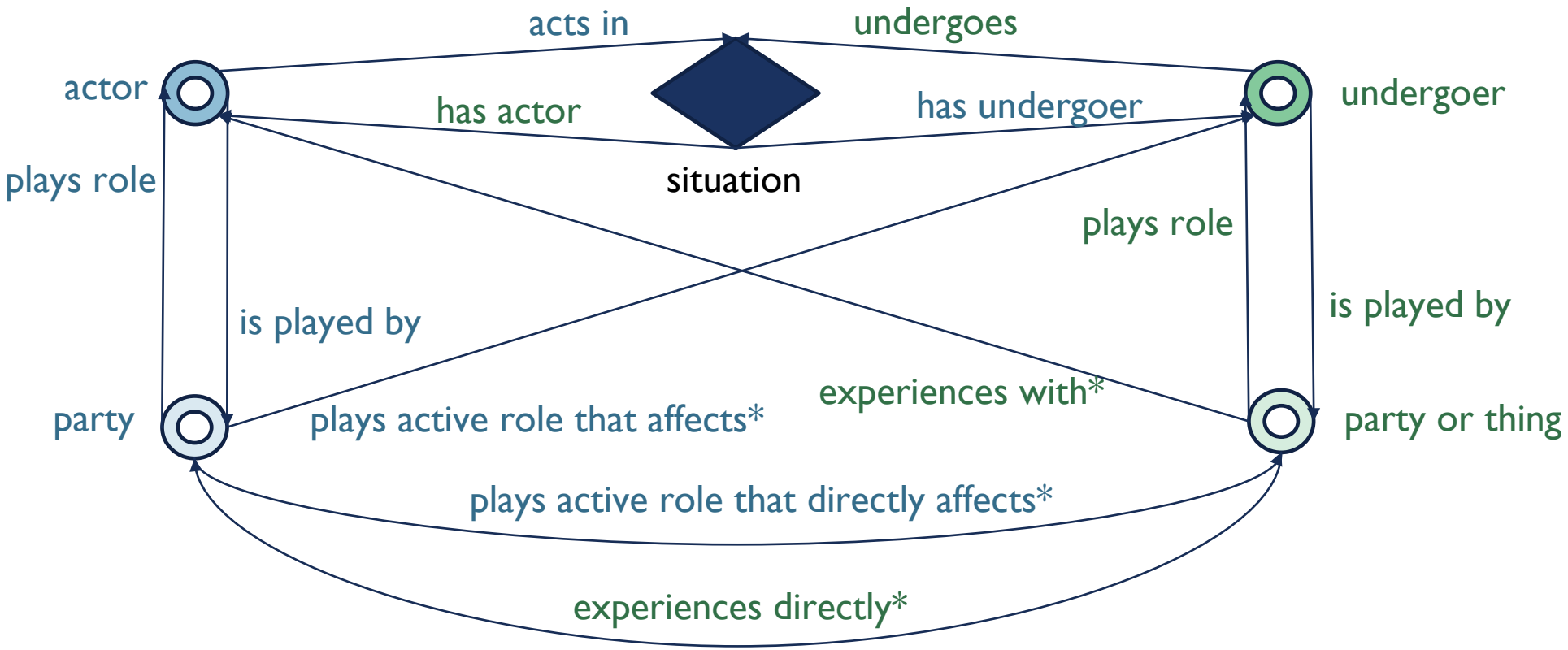
* Properties are chained; each pair of chained properties has a corresponding party to role inverse

Situation with Role to Role/Party relations



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

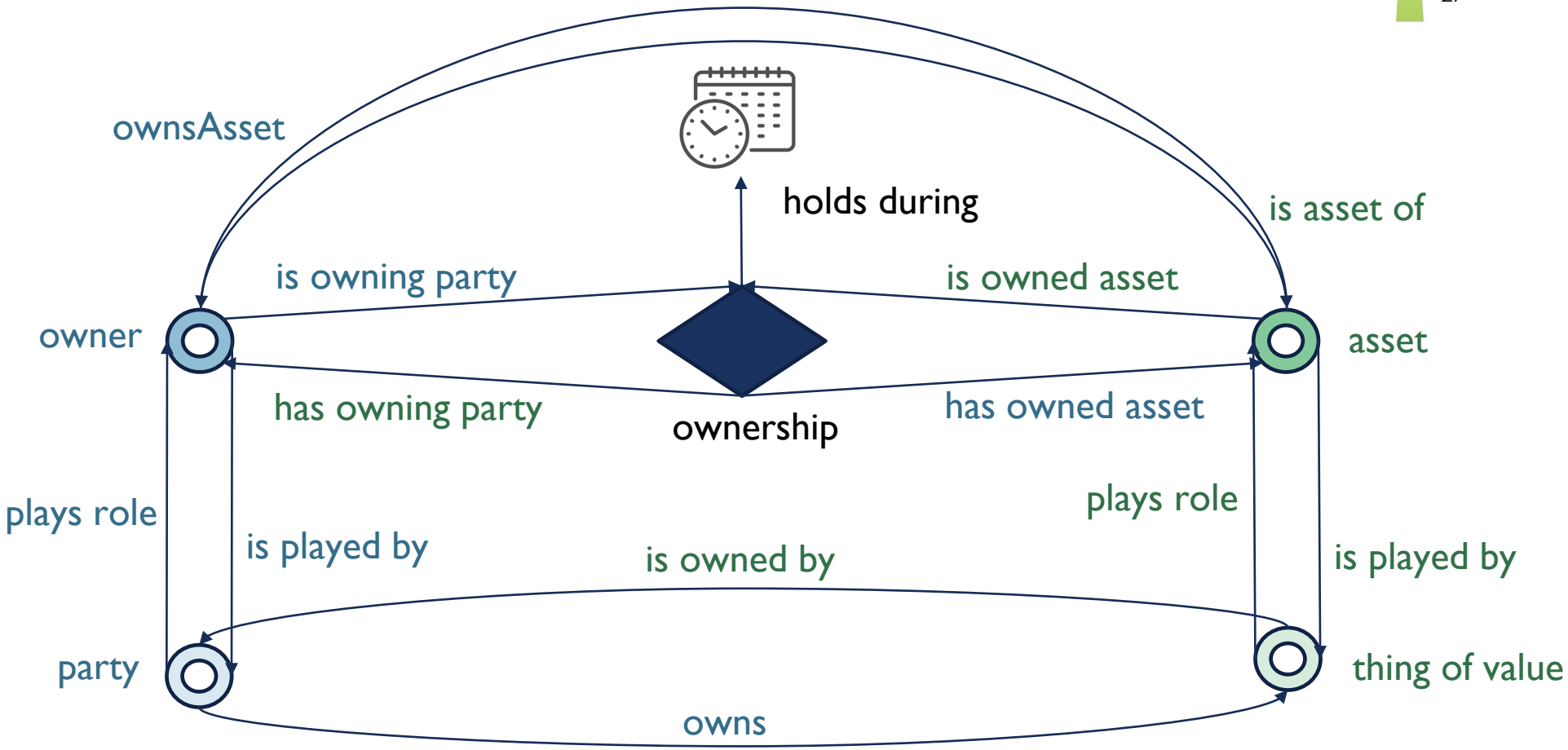
Situation with Party to Party/Role relations



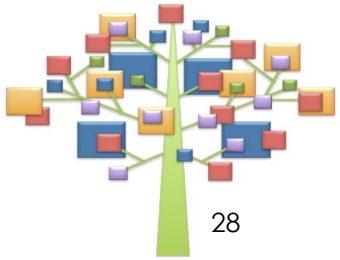
* Properties are chained; Each pair of chained properties has a corresponding party to role inverse

Why so complicated?

- Ownership may be partial / shared
- 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



Next steps



- Additional ontologies that are planned for a Commons 1.2 revision
  - Tensors, Vectors and Arrays
  - Tensor and vector quantities, and transformation of the SysML v2 library of over 700 units
  - Products and Services, including support for GS1 product identifiers (GTIN, UPC)
  - Organizations and Legal Entities
  - Registration Authorities
  - Regulatory Agencies
  - …
- Target is Q2 2024 June meeting, but that may slip depending on requirements
- Commons Ontology Library is available at – https://www.omg.org/spec/COMMONS
- Multiple Vocabulary Facility (MVF) is available at – https://www.omg.org/spec/MVF