PROV-O

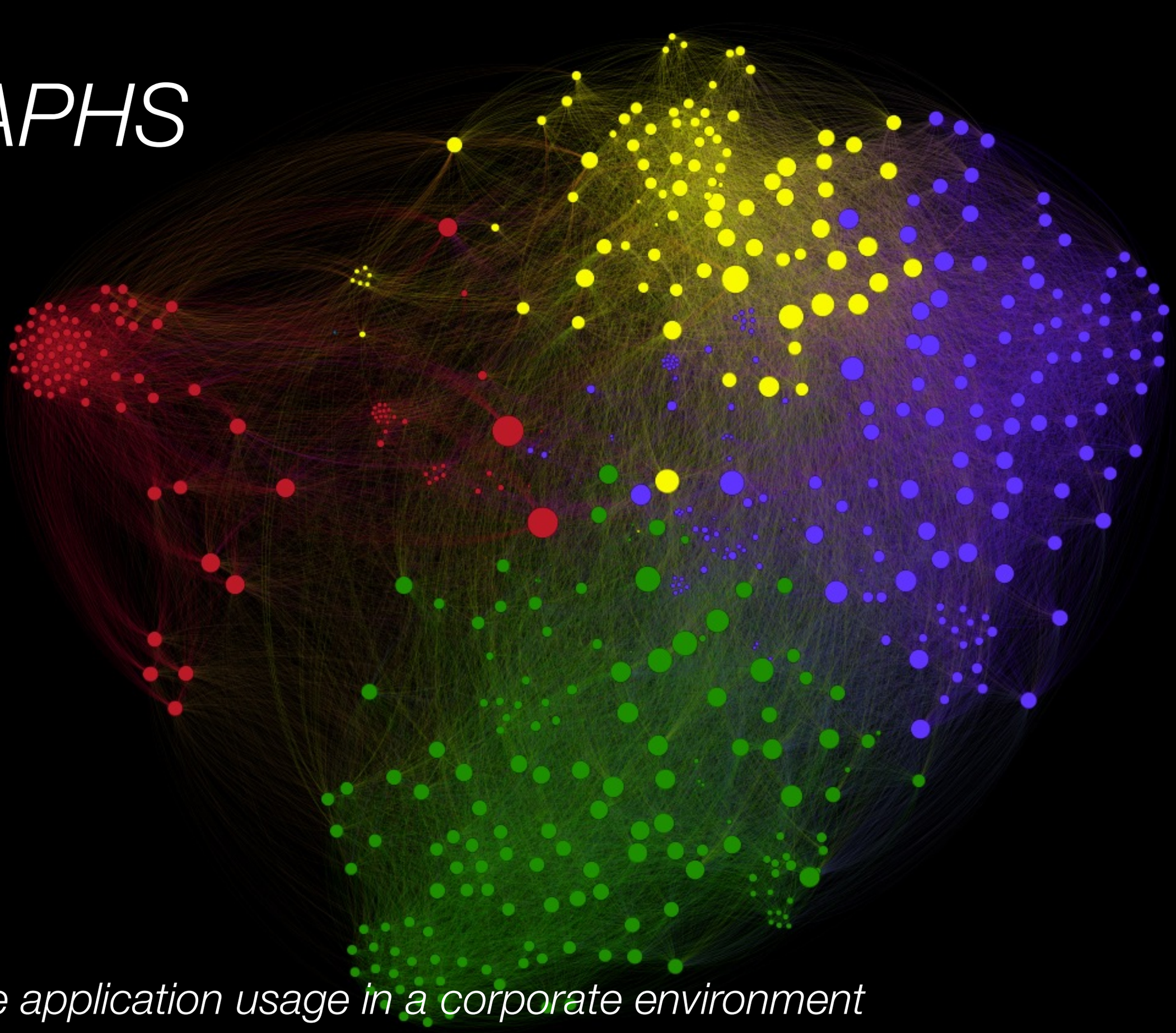
OWL

RDF

Graphs

BIOREGIONAL
ASSESSMENTS

GRAPHS



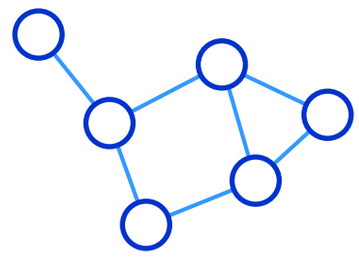
Software application usage in a corporate environment

Created using NetworkX & GEPHI

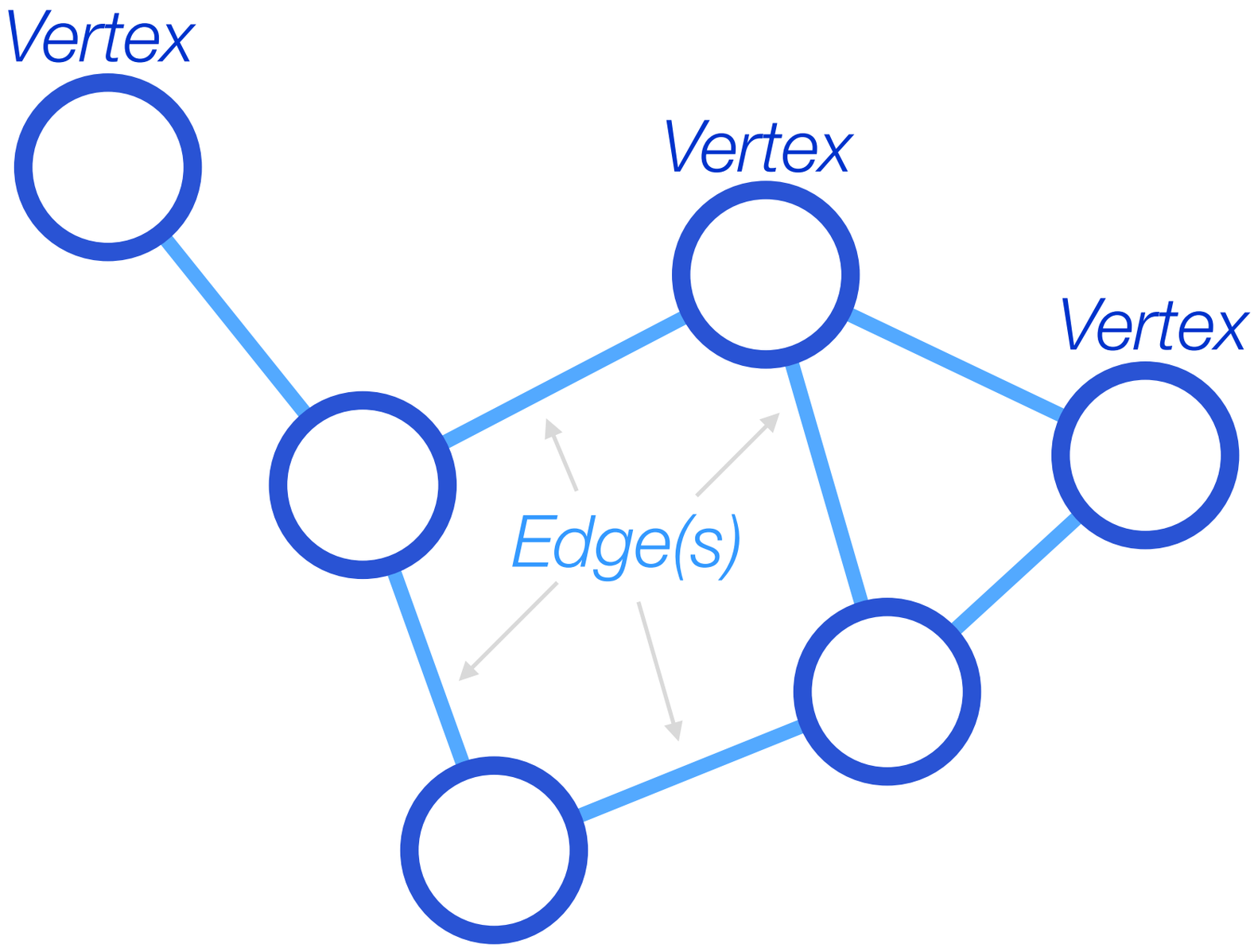
WHAT'S A GRAPH?

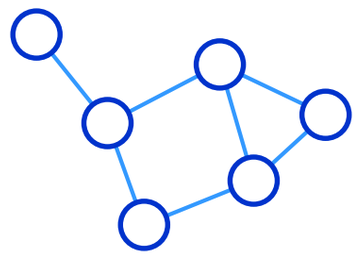
Representation of a set of objects where ***some pairs*** of objects are connected by links.

Links are binary. One link connects two *(and only two)* objects.



GRAPH TERMINOLOGY

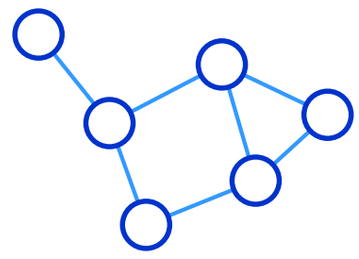




TERMINOLOGY DIFFERENCES

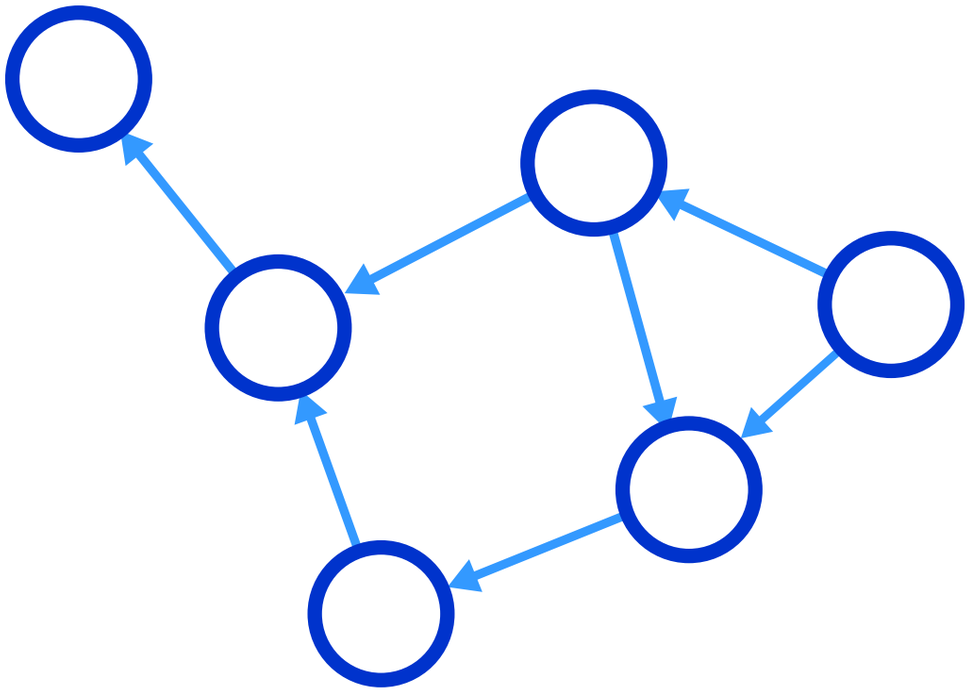
Network Science and *Graph Theory* use different words for the same concepts.

Network Science	Graph Theory
Network	Graph
Node	Vertex
Link	Edge

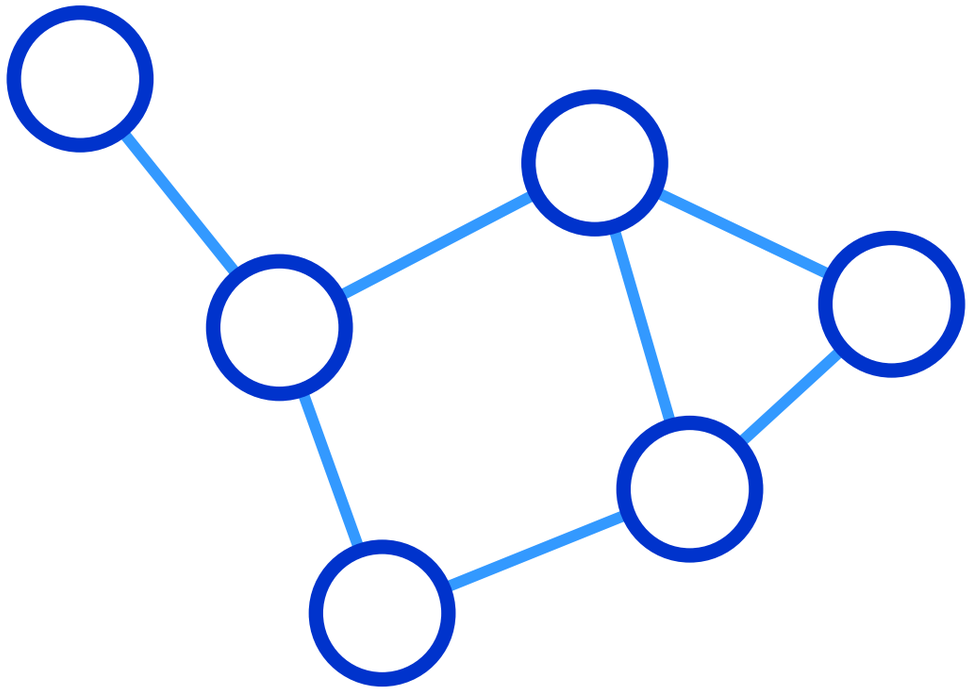


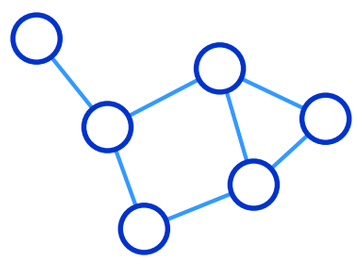
GRAPH TYPES

Directed



Undirected





GRAPH SUB TYPES

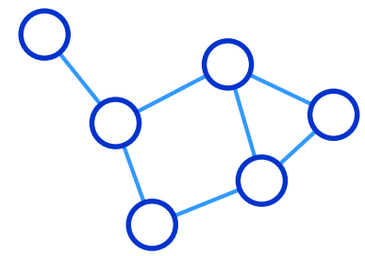
Complete

$$\textit{Number of edges} = \frac{n\,(n-1)}{2}$$

Bipartite

Weighted

Subgraph



GRAPH USES

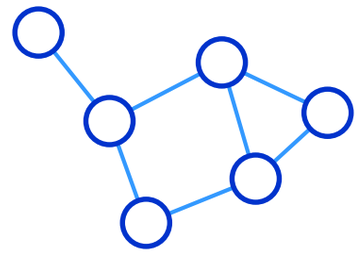
Representation

As an alternative to:

- Inverted trees (XML)
- Lists (flat files)
- Relations (RDMS).

Analysis & Prediction

- Degree
- Betweenness
- Shortest path
- Components
- Pagerank




RDF

WHO USES RDF?

www.data.gov.au for metadata

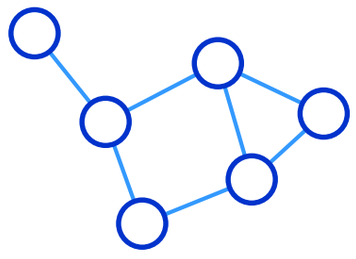
- **Dublin Core** (Metadata)
- **AGLS** (Australian Government Locator Service). Extension of Dublin Core
- **Data Catalog Vocabulary** (DCAT) (Describes Data Sets)

Bureau of Meteorology (Australia)

ACORN-SAT Linked Climate Dataset [http://lab.environment.data.gov.au] Transformation of tabular data into a Linked Sensor Data Cub based on the W3C Semantic Sensor Network Ontology,

See related paper here: http://www.semantic-web-journal.net/system/files/swj457.pdf. (Authors Andrew Wolfe et al). Project won a national innovation award in 2013. (see IAWARDS)

Data.gov.au was used as the root domain for URI sets. SPARQL was used as the query language.



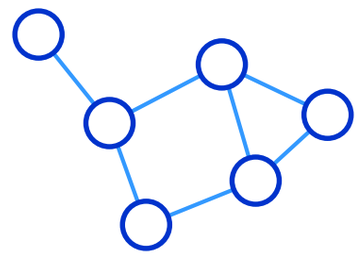
WHAT IS RDF?

Resource Description Format

Origins as a metadata format

Now, part of W3C's semantic web stack.

Represents graphs; Basic unit is the *'triple'*



RDF's BASIC CONSTRUCT: THE TRIPLE

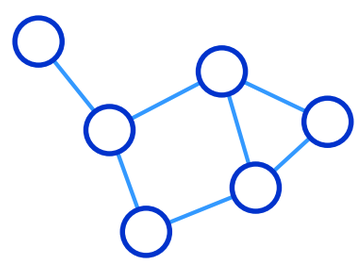
Subject - Predicate - Object

Subject

Object

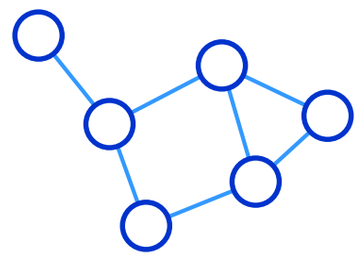
"Andrew plays Rugby"

Predicate



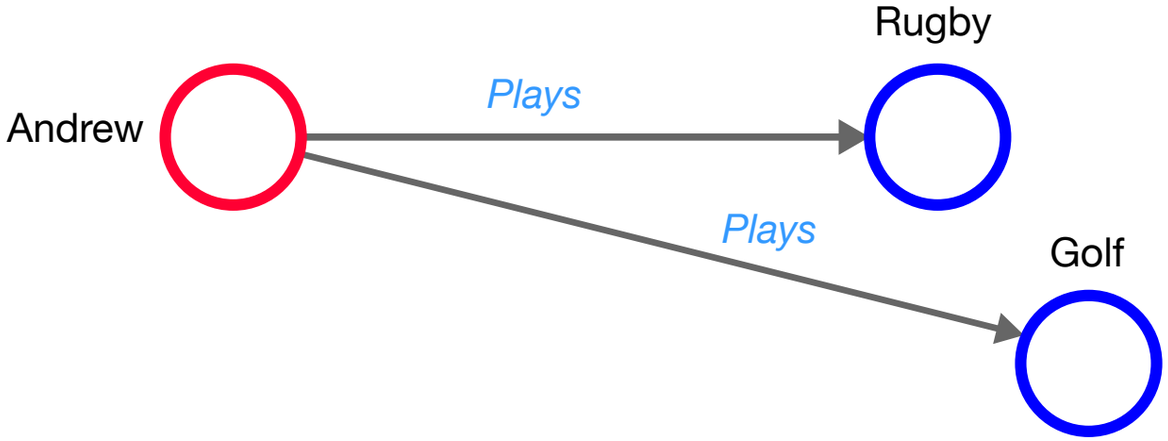
THE TRIPLE AS A GRAPH (1)

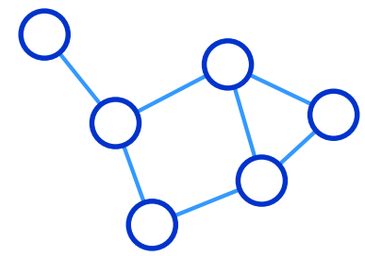
Easy !



THE TRIPLE AS A GRAPH (2)

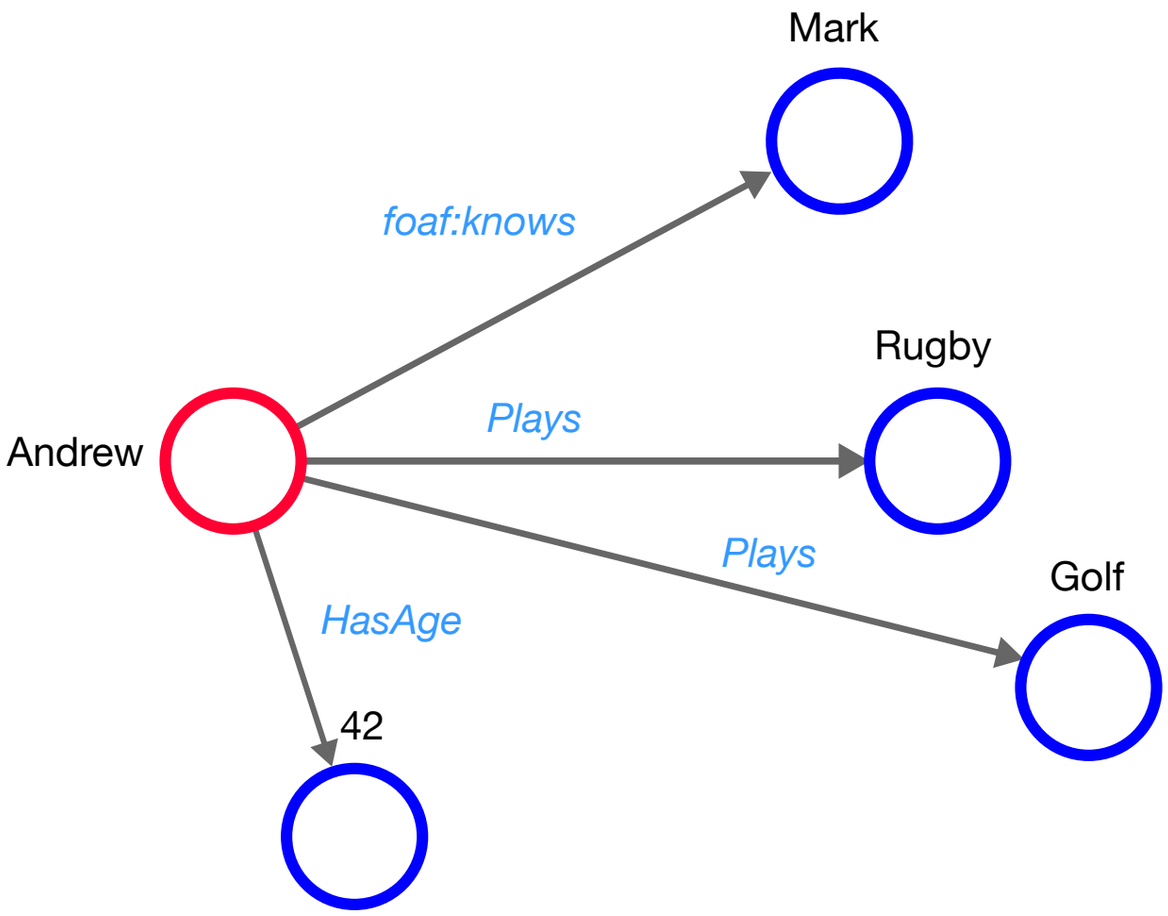
There’s more to Andrew than Rugby!

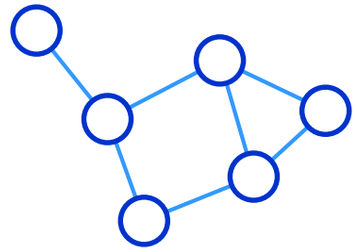




THE TRIPLE AS A GRAPH (3)

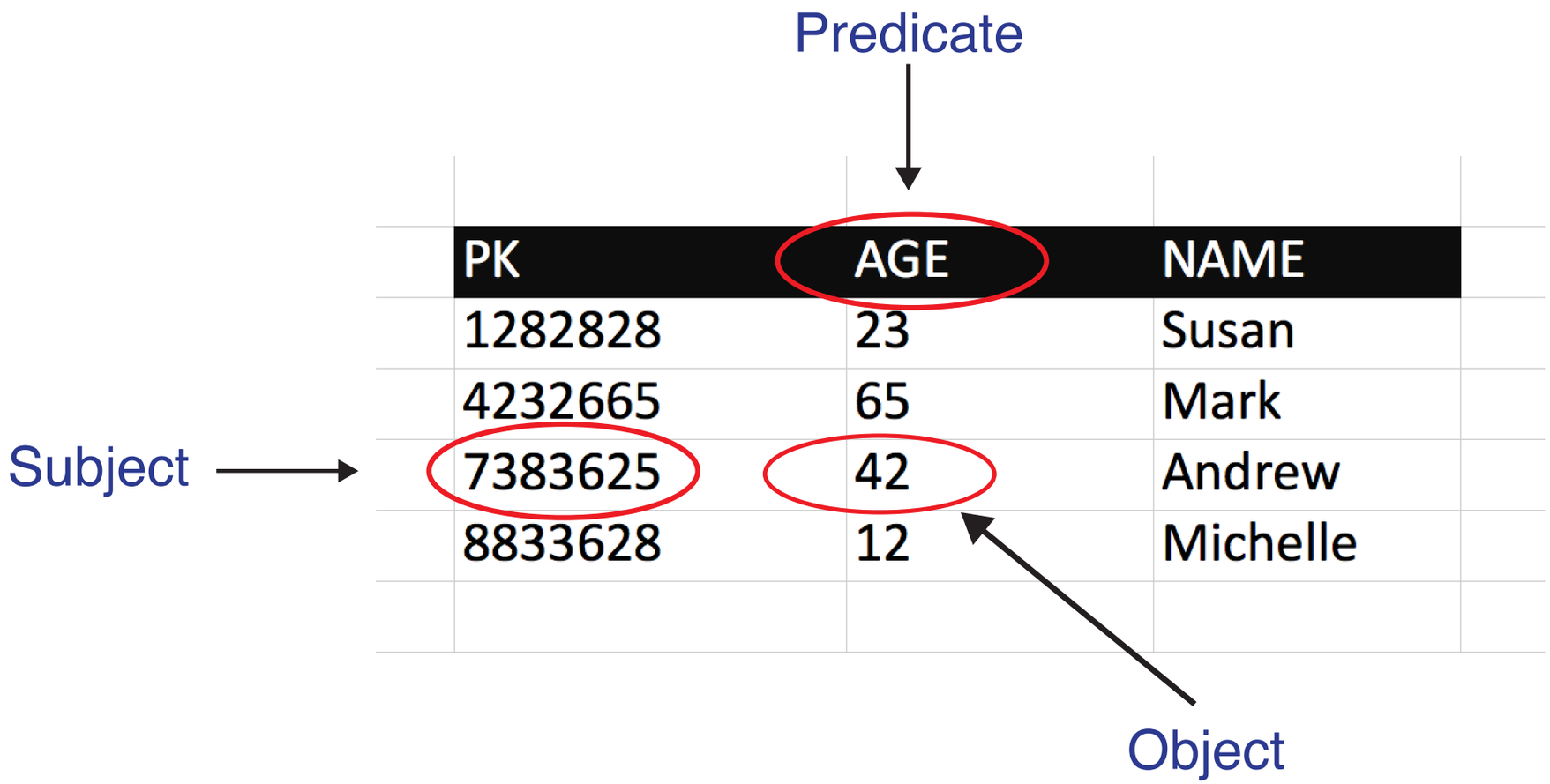
Add two more activities

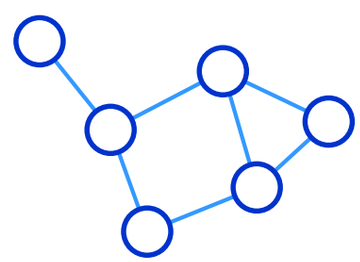




GRAPHS VERSUS DATABASES (1)

Graphs and relations have similarities





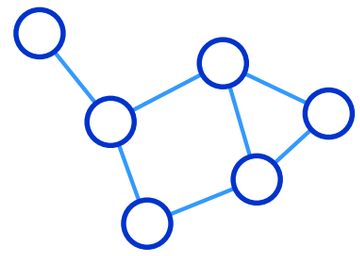
GRAPHS VERSUS DATABASES (2)

Advantages of graphs

- **Do not require joins -** they can scale for very large datasets.
- **Flexibility** - Schema does not need to be defined in advance.
- **Different structures -** Can query and perform different analysis techniques.

Advantages of databases

- **Depends on underlying structure -** Better at tabular data
- **Same operation** - Better at performing same operation on large sets of homogenous data. Sum of a million numbers.



RDF SERIALISATION (1) - XML

The ‘knows’ predicate comes from the Friend of a Friend (FOAF) namespace

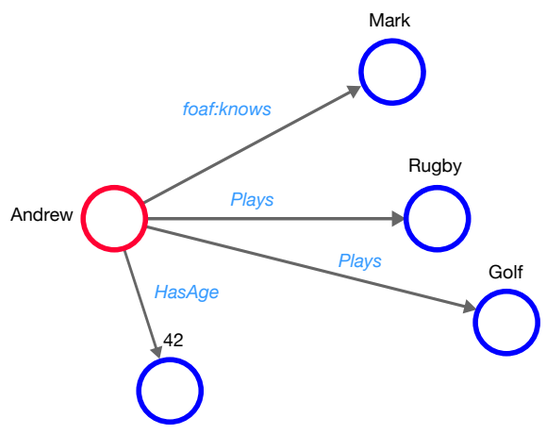
The 42 object of hasAge is an XSD integer.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">]>
<rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:terms="http://exampleterms.com/terms/"
        xmlns:foaf="http://xmlns.com/foaf/0.1/">
            <rdf:Description rdf:about="www.andrew.com/people#andrew">
                    <terms:plays>rugby</terms:plays>
                    <terms:plays>golf</terms:plays>
                    <terms:hasAge rdf:datatype="&xsd;integer">42</terms:hasAge>
                    <foaf:knows> rdf:resource="www.andrew.com/people#mark"</foaf:knows>
            </rdf:Description>
</rdf:RDF>
```

The predicates must refer to a namespace or URI. To accomodate this requirement, the ‘terms’ namespace refers to a generic location. The rdf:Description element encloses all predicates for the subject focussed triple set.

The rdf:Description elements defines the subject. The rdf:about property points to an URI reference. Subjects of statements are referred to using the rdf:about tag and objects of predictates use the rdf:resource tag.

Original triples



RDF SERIALISATION (2) - TURTLE

```
andrew
    a person;
    plays golf, rugby;
    hasAge 42;
    foaf:knows mark.
```

Annotated example with prefixes defined

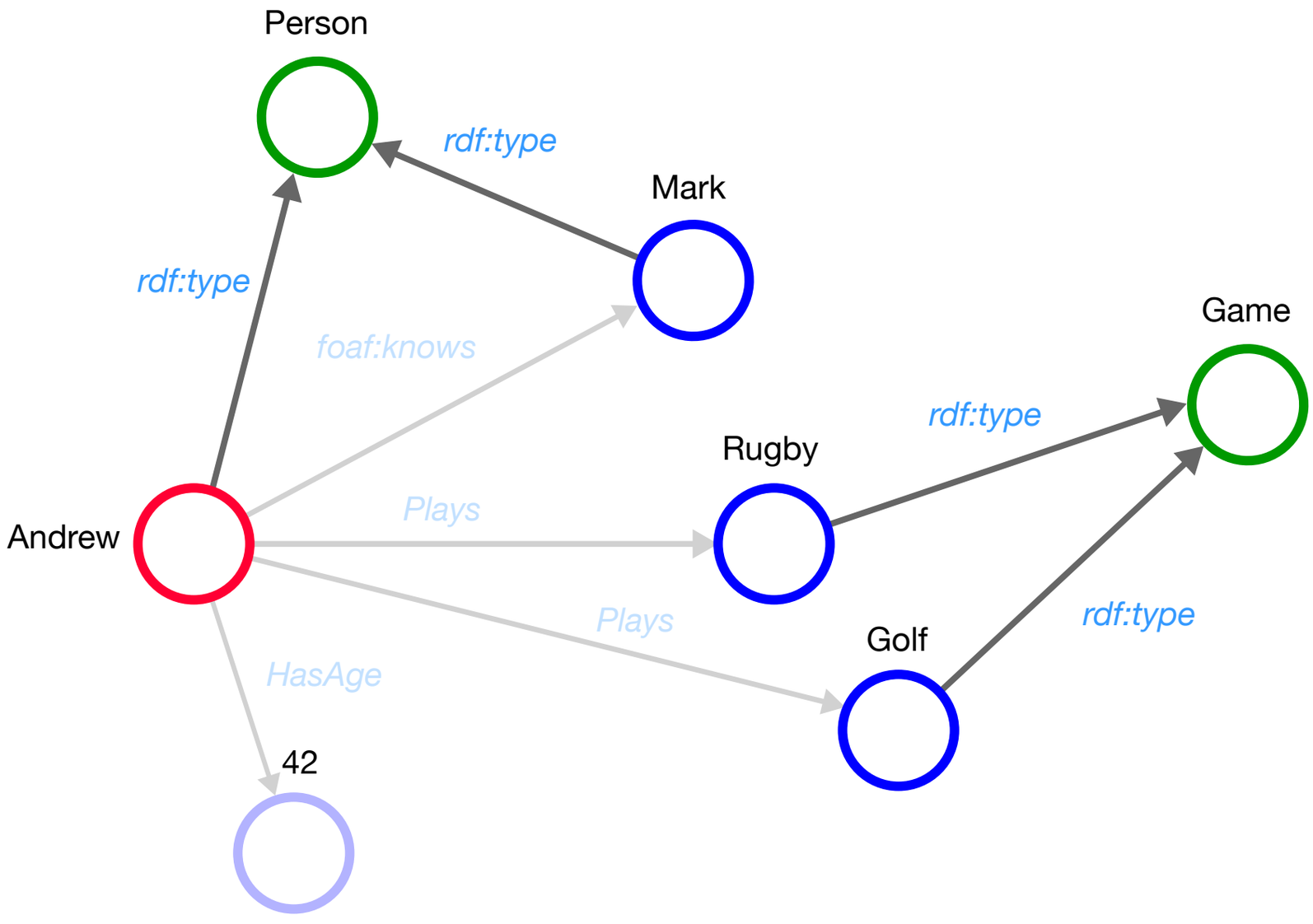
```
@prefix : <www.andrew.com/people#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix terms: <http://exampleterms.com/terms/> .

    :andrew
            terms:plays "golf", "rugby";
            terms:hasAge 42;
            foaf:knows :mark.
```

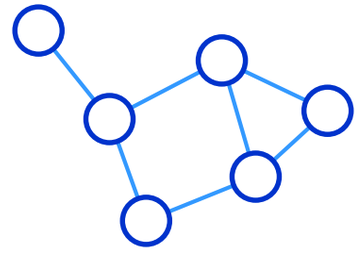
Commas separate objects that refer to the same predicate. Semi-colons separate predicates that refer to the same subject.

A full stop terminates all expressions that refer to the same subject.

RDF EXTENSION (1) - CLASSES

Use the *rdf:type* property

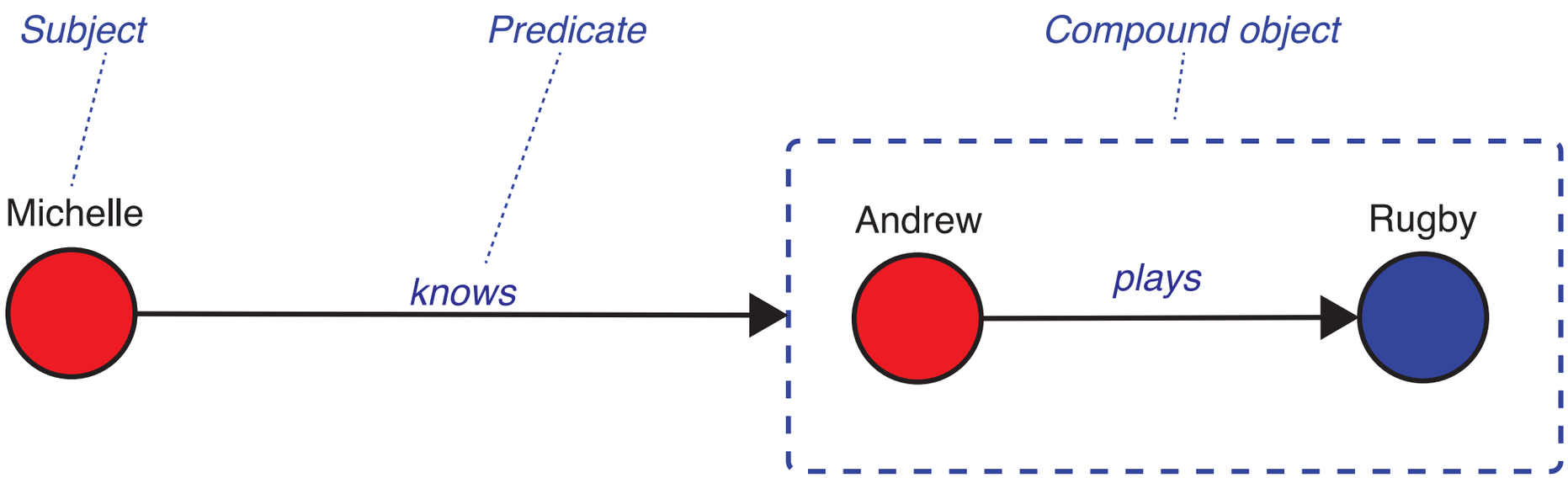




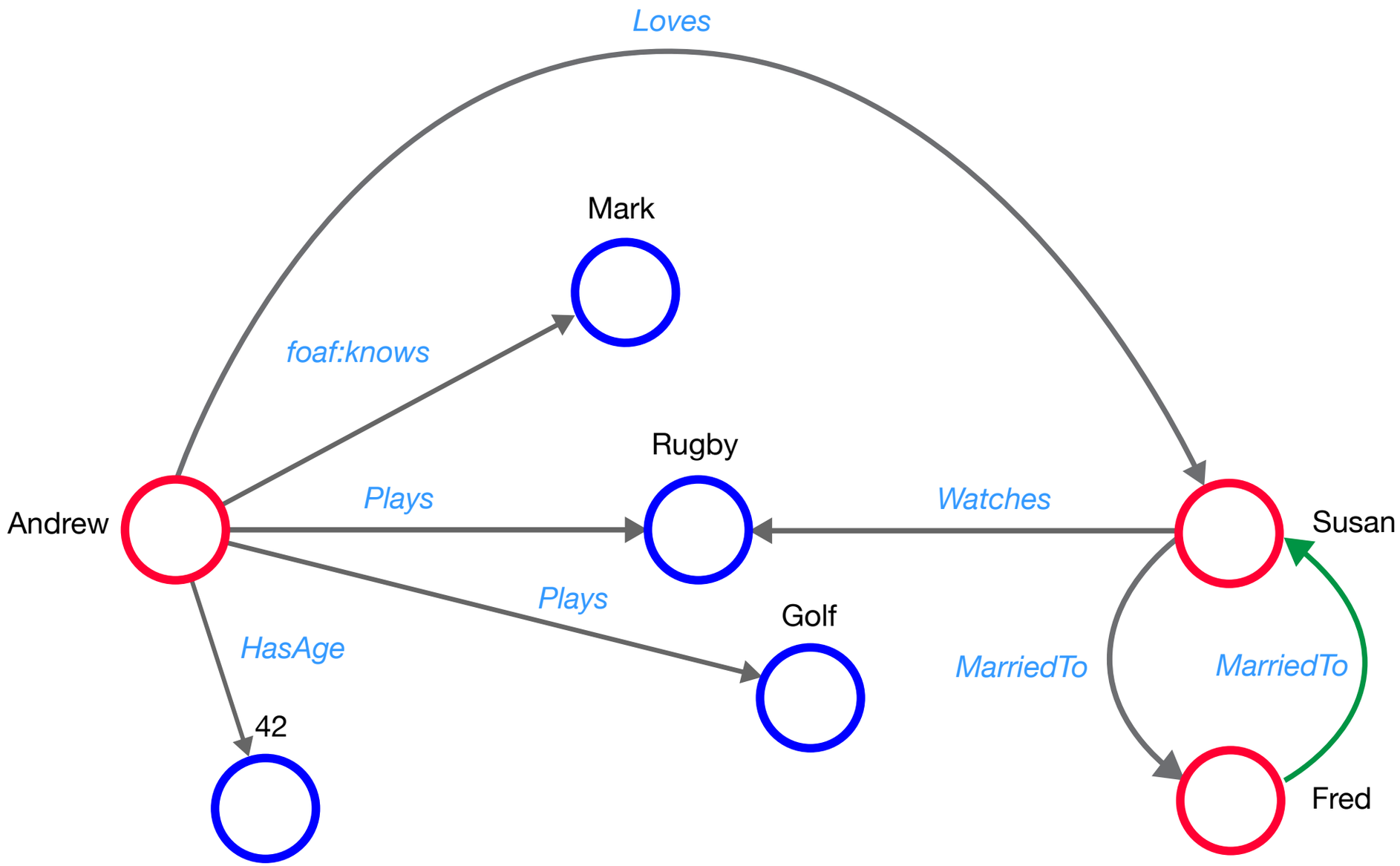
RDF EXTENSION (2) - REIFICATION

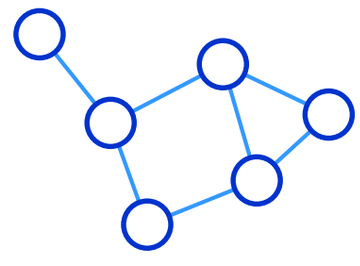
A difficult word but an easy concept

Statement about another triple



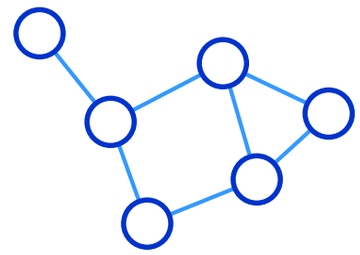
MORE COMPREHENSIVE EXAMPLE





INTRODUCTION TO REASONING (1.A)

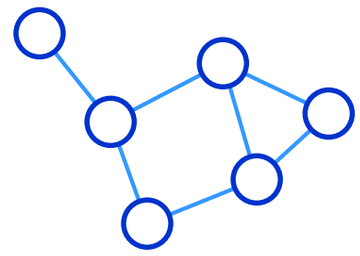
MarriedTo(Susan, Fred) *assertion 1*



INTRODUCTION TO REASONING (1.B)

MarriedTo(Susan, Fred) *assertion 1*

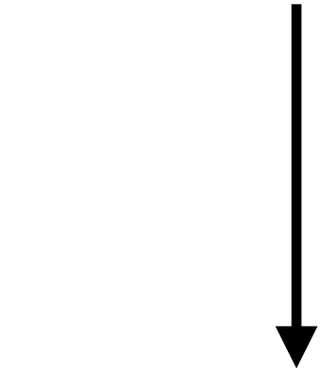
Symmetrical(MarriedTo) *assertion 2*



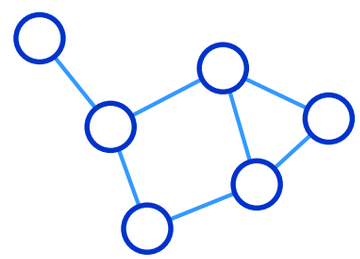
INTRODUCTION TO REASONING (1.C)

MarriedTo(Susan, Fred) *assertion 1*

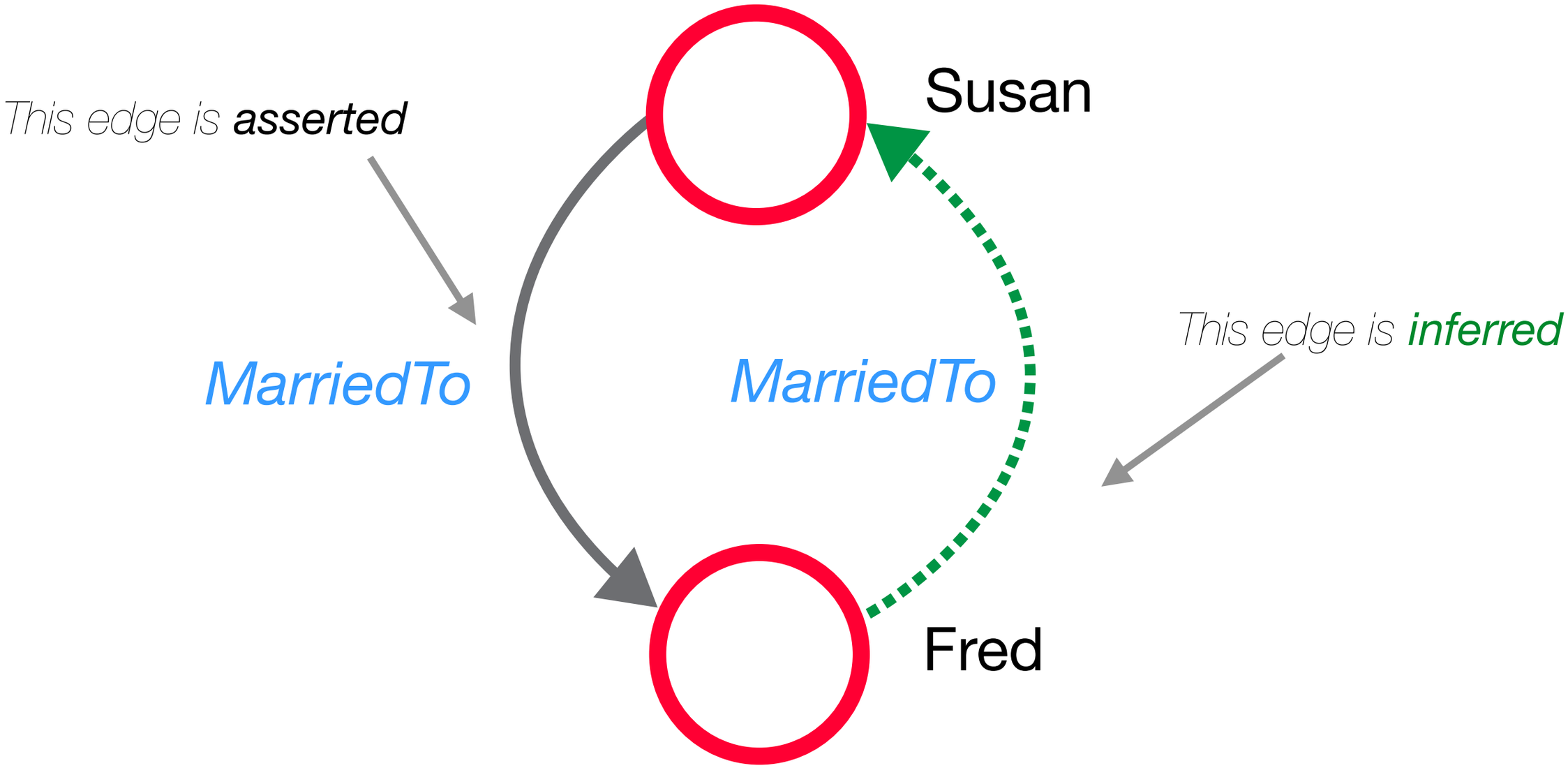
Symmetrical(MarriedTo) *assertion 2*



MarriedTo(Fred, Susan) *entailment*

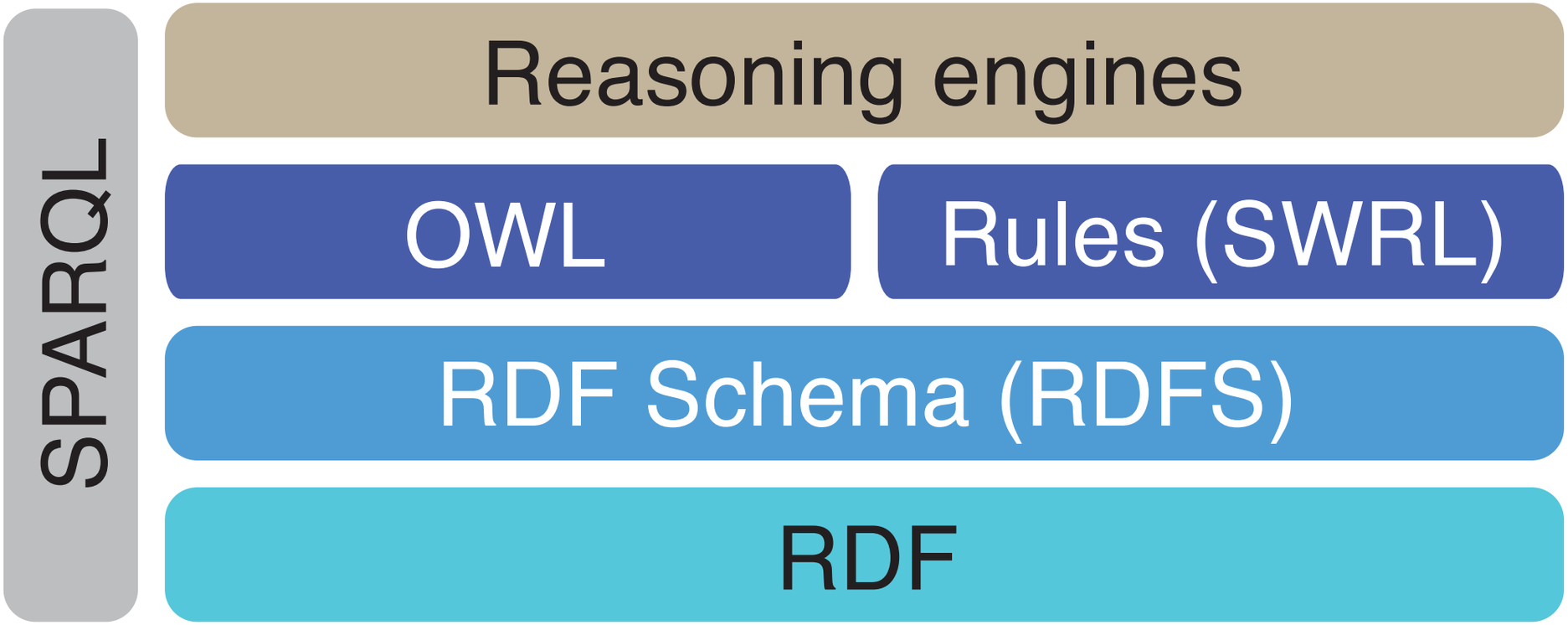


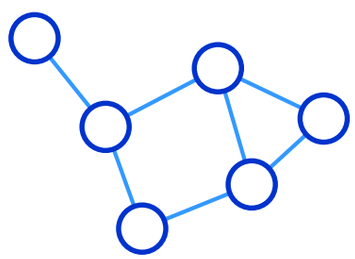
INTRODUCTION TO REASONING (1.D)




Semantic
Web

THE SEMANTIC WEB LAYER CAKE





THE SEMANTIC WEB & PROV-O

PROV-O

SPARQL

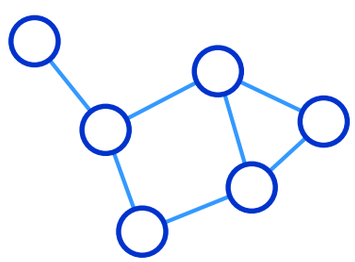
Reasoning engines

OWL

Rules (SWRL)

RDF Schema (RDFS)

RDF

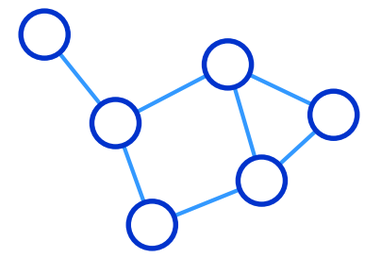


VOCABULARIES, ONTOLOGIES…ETC.

Vocabulary - collection of classes & predicates published in regards to a specific theme.

Example - Friend of a Friend *(FOAF project)*
www.foaf-project.org

Many Other examples - Dublin Core & DCAT…



FOAF EXAMPLE TERMS

Person

name

title

familyName

givenName

knows

age

img

nick

mbox

homepage

weblog

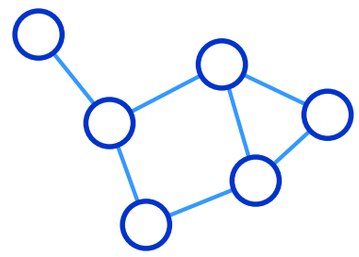
jabberID

mbox

interest

publications

logo



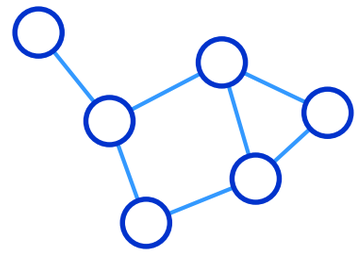
WHAT'S AN ONTOLOGY? [1]

Computer Science Perspective - "Finite list of terms and associated relationships…a formal specification of a concept" *(Antoniou G & Harmelen F)*

WHAT'S AN ONTOLOGY? [2]

Difference to a vocabulary - More expressive. An ontology has comprehensive set of rules and can generate inferences.

Requires a more expressive formalism than compared to RDF.



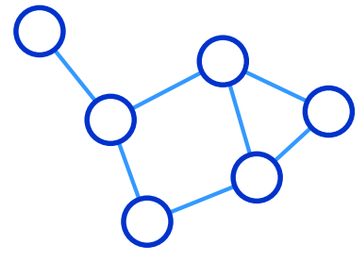
RDFS…SLIGHTLY MORE EXPRESSIVE

Extra classes, extra properties - RDFS defines an extra 4 classes and 6 properties.

Adds minimal reasoning - RDFS adds simple reasoning capabilities and some annotation ability.

REASONING WITH RDFS

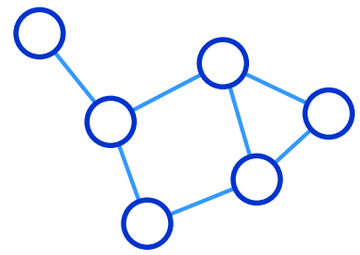
FreelancesTo(David, ABC Ltd) *assertion 1*



REASONING WITH RDFS

FreelancesTo(David, ABC Ltd) *assertion 1*

rdfs:subPropertyOf(FreelancesTo, WorksFor) *assertion 2*

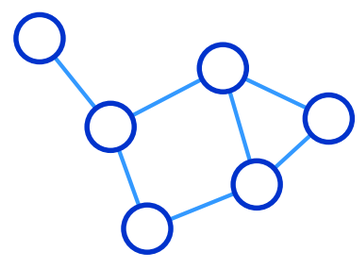


REASONING WITH RDFS

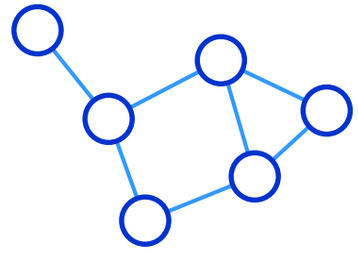
FreelancesTo(David, ABC Ltd) *assertion 1*

rdfs:subPropertyOf(FreelancesTo, WorksFor) *assertion 2*

WorksFor(David, ABC Ltd) *entailment*



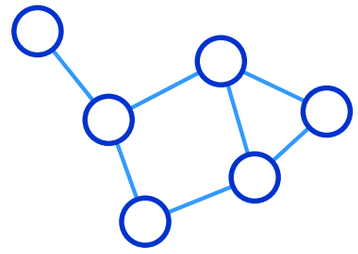
REASONING WITH RDFS



Ontology Web Language (OWL)

REASONING WITH OWL [1]

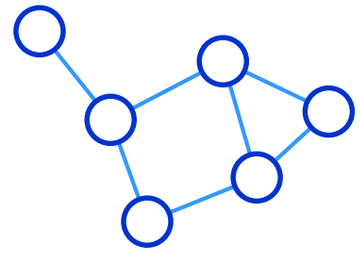
DependsOn(Report, Database) *assertion 1*



REASONING WITH OWL [2]

DependsOn(Report, Database) *assertion 1*

owl:InverseOf(DependsOn, Enables) *assertion 2*

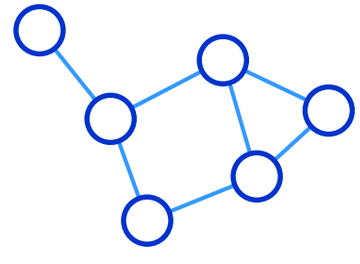


REASONING WITH OWL [3]

DependsOn(Report, Database) *assertion 1*

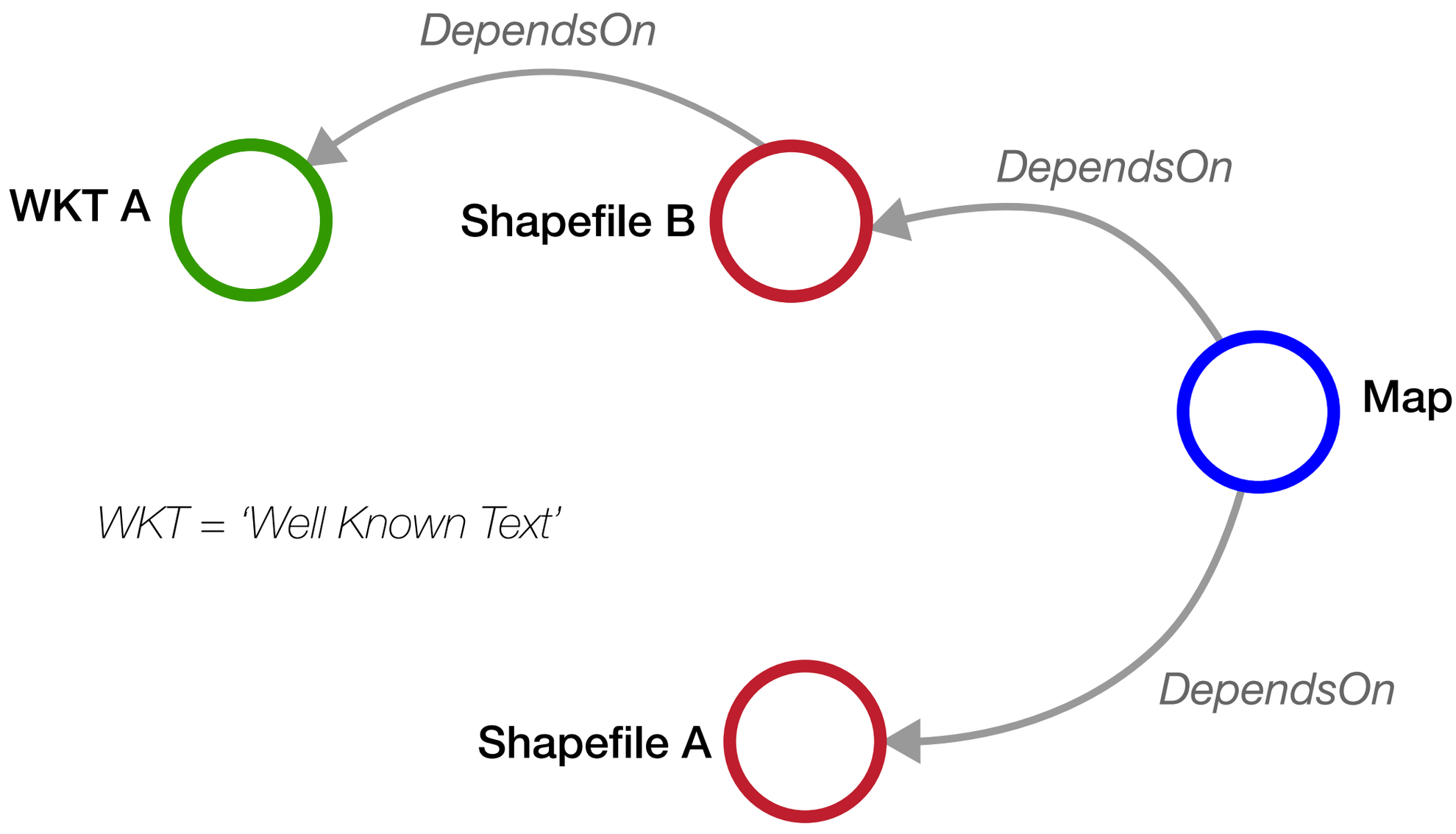
owl:InverseOf(DependsOn, Enables) *assertion 2*

Enables(Database, Report) *entailment*



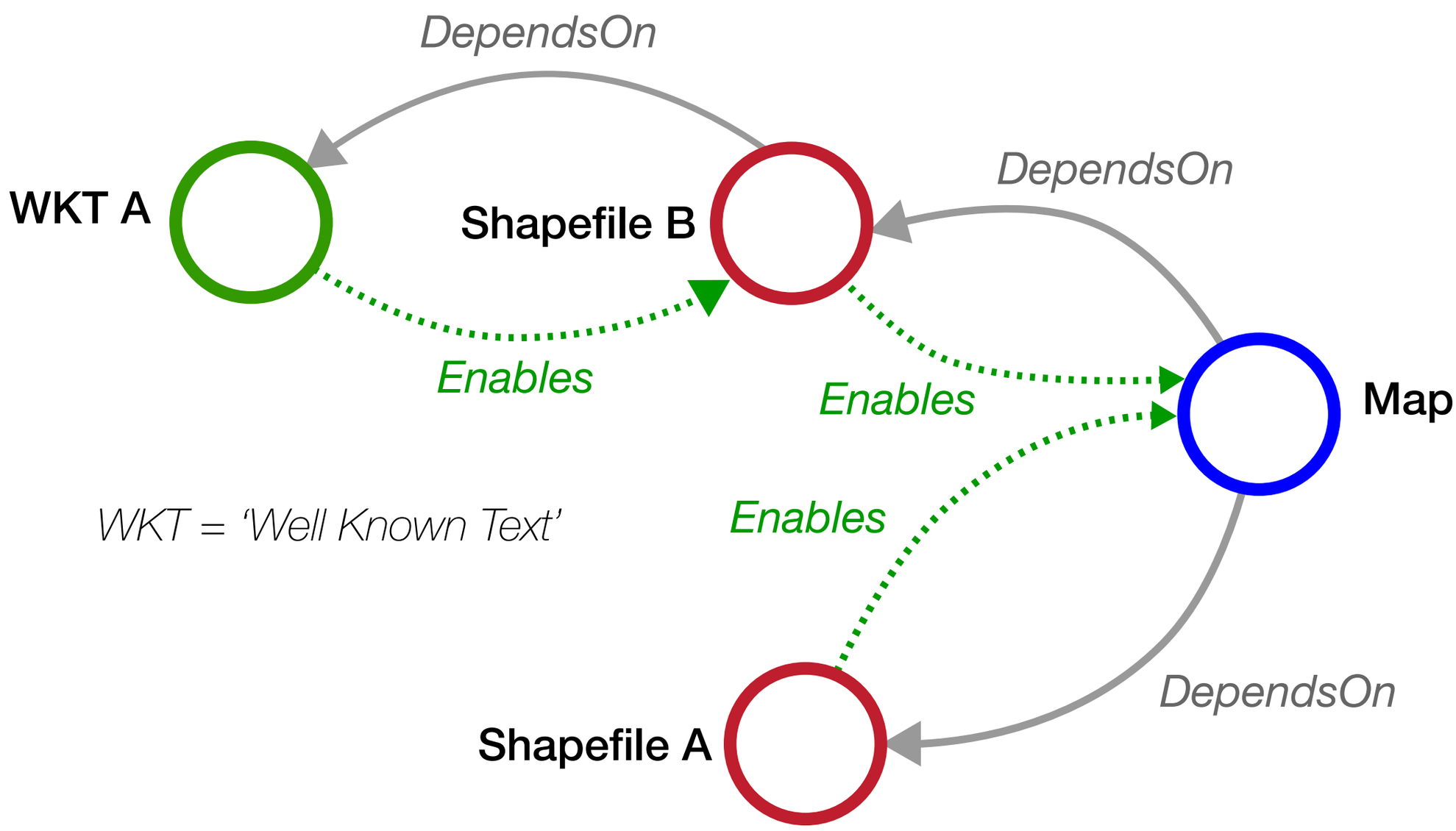
REASONING WITH OWL [4]

Before reasoning



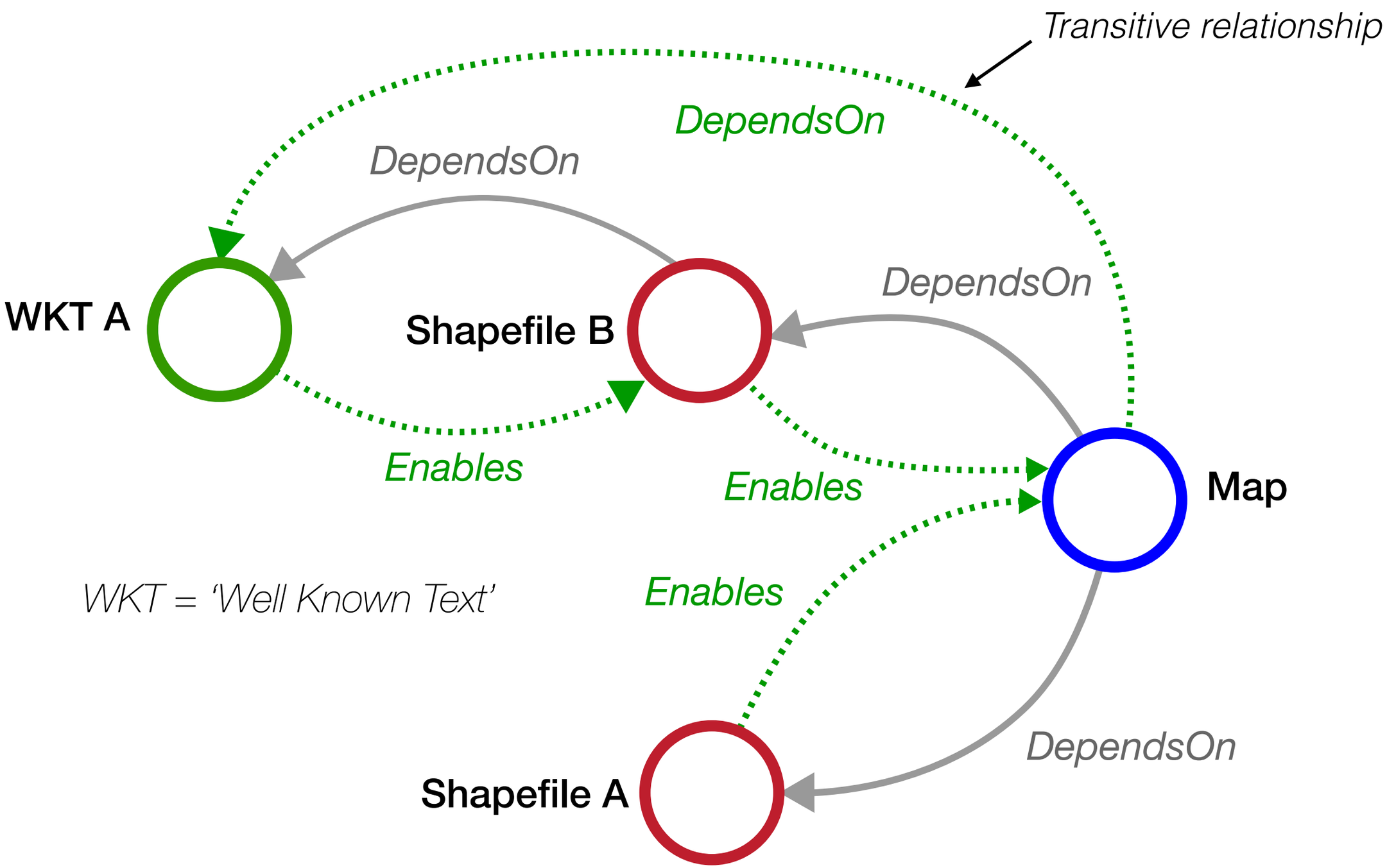
REASONING WITH OWL [5]

After reasoning



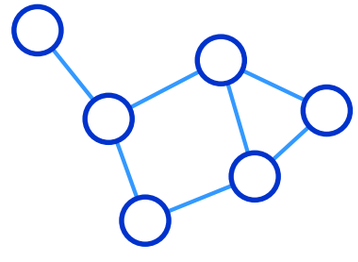
WKT = 'Well Known Text'

REASONING WITH OWL [6]



WKT = 'Well Known Text'

PROV-O



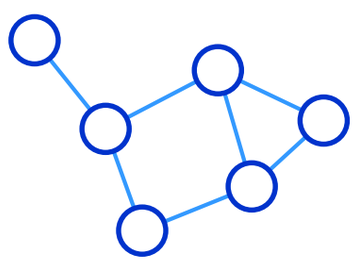
PROV-O (PROVENANCE ONTOLOGY)

W3C Recommendation

3 Different Levels *(Starting, Expanded & Qualified)*

reference: http://www.w3.org/TR/prov-o/

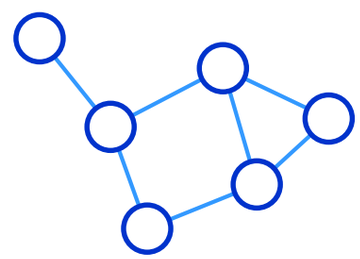
***'Simple'* is starting point** *(3 classes, 9 properties)*



WHAT DOES PROVENANCE MEAN? (1)

"Place of origin or earliest known history of something: an orange rug of Iranian provenance."

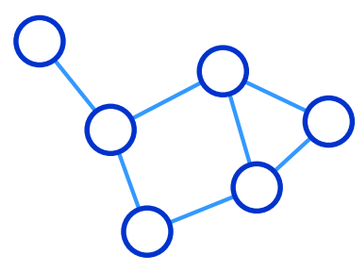
Source: Oxford Dictionary of English



WHAT DOES PROVENANCE MEAN? (2)

"Provenance is defined as a record that describes the people, institutions, entities and activities involved in producing, influencing or delivering a piece of data or a thing."

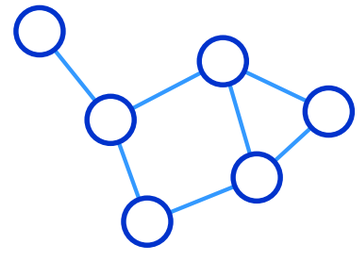
Source: WC3 http://www.w3.org/2005/Incubator/prov/wiki/What_Is_Provenance



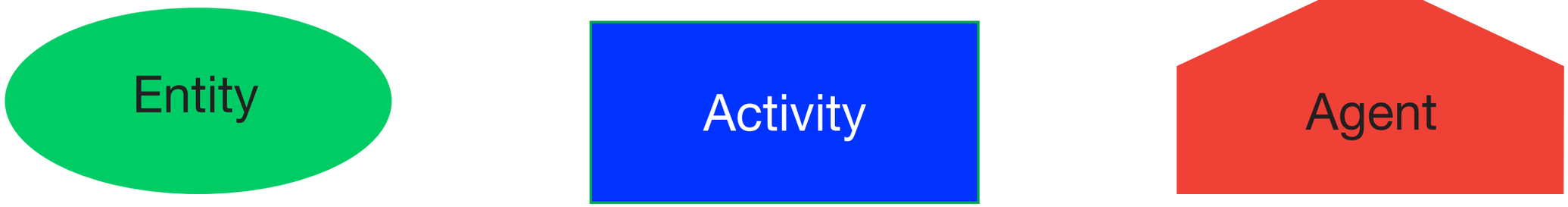
DIFFERENCE TO METADATA

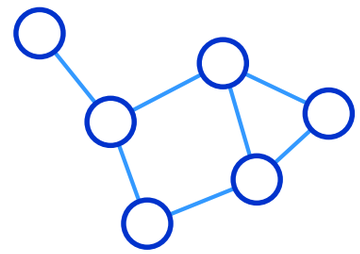
"Provenance assertions are a form of contextual metadata and can themselves become important records with their own provenance."

Source: W3C http://www.w3.org/TR/prov-dm/

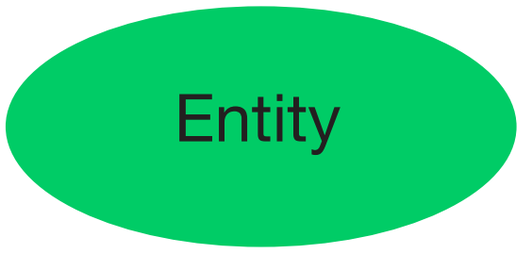


PROV-O SIMPLE - THREE CLASSES





PROV-O FORMAL DEFINITIONS

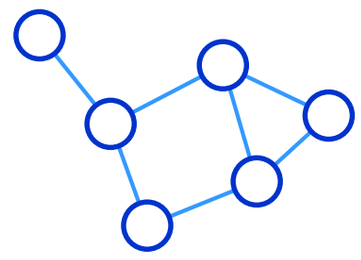


“An *entity* is a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary”.

“An *activity* is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using or generating entities”.

“An *agent* is something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent’s activity”

Source: W3C®



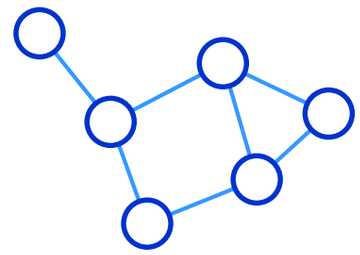
PROV-O EXAMPLE (1)

Dataset

Postcode	Sales
2600	5
2601	3
2602	5
2603	2
2604	6
2605	8
2600	2
2601	4
2602	2
2603	1
2604	5

Post Codes

Postcode	Name
2600	Deakin
2604	Kingston
2603	Manuka



PROV-O EXAMPLE (2)

Dataset

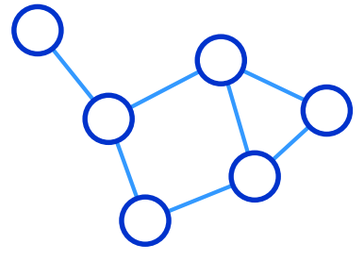
Postcode	Sales
2600	5
2601	3
2602	5
2603	2
2604	6
2605	8
2600	2
2601	4
2602	2
2603	1
2604	5

Post Codes

Postcode	Name
2600	Deakin
2604	Kingston
2603	Manuka

Composition

Suburb	Total Sales
Deakin	7
Kingston	3
Manuka	11

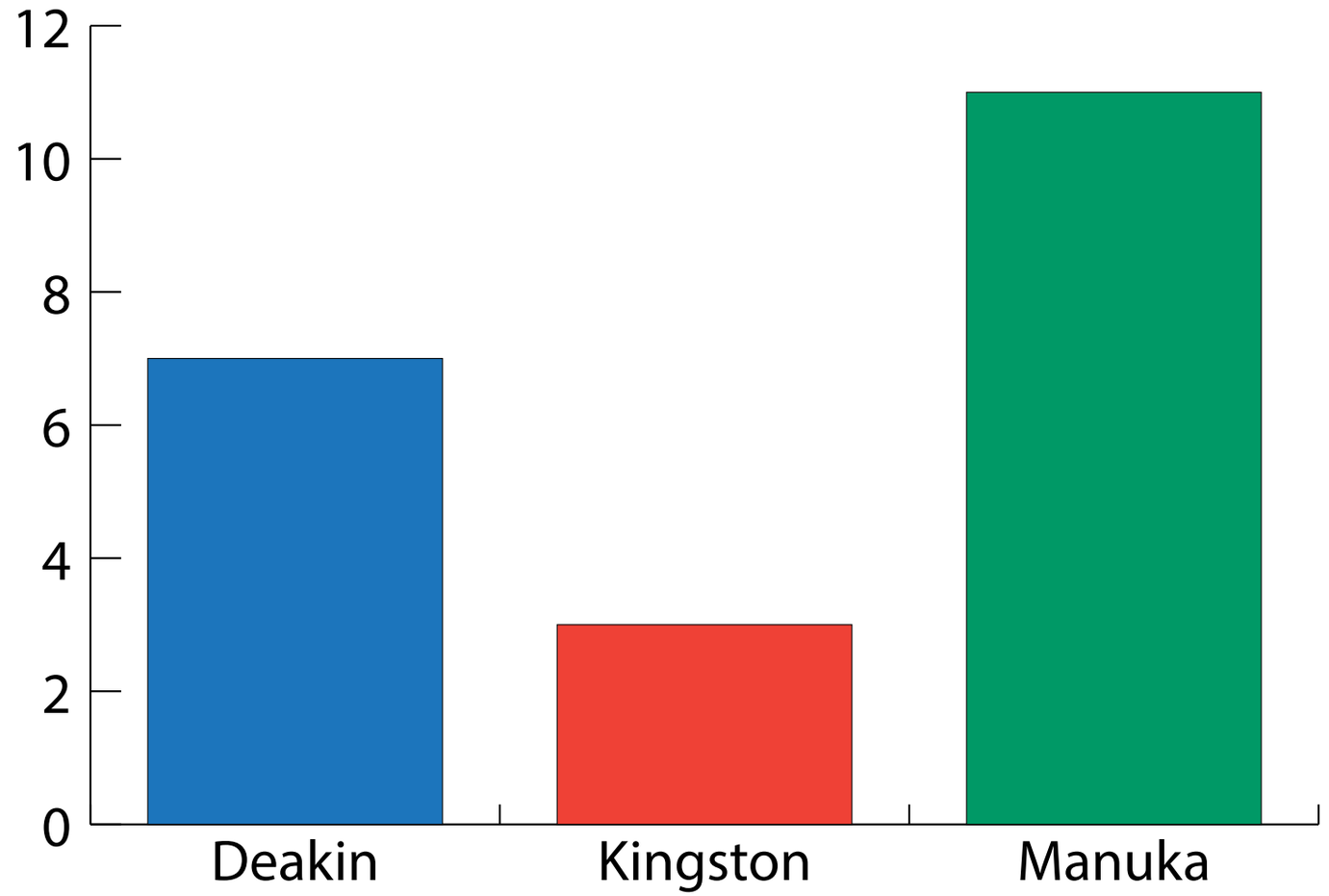


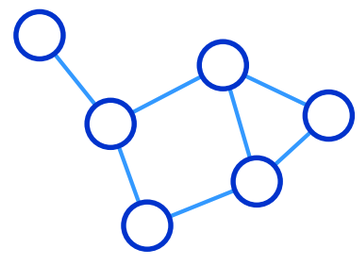
PROV-O EXAMPLE (3)

Aggregation

Suburb	Total Sales
Deakin	7
Kingston	3
Manuka	11

Graph

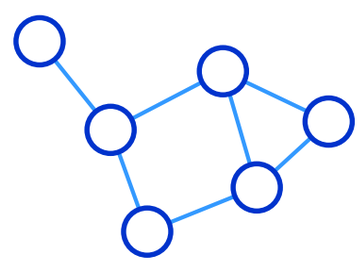




PROV-O EXAMPLE (4)

Focused on history

Arrows face from the future to the past



PROV-O EXAMPLE (5)

ex:rawData

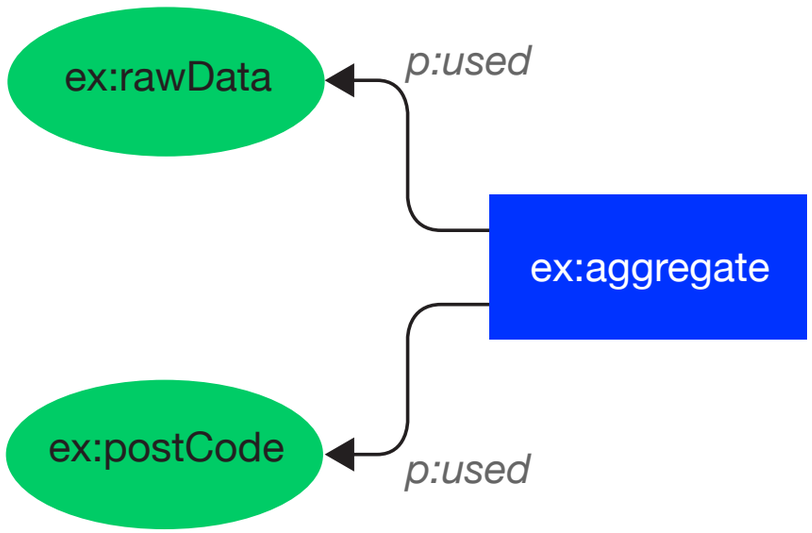
ex:postCode

Activity

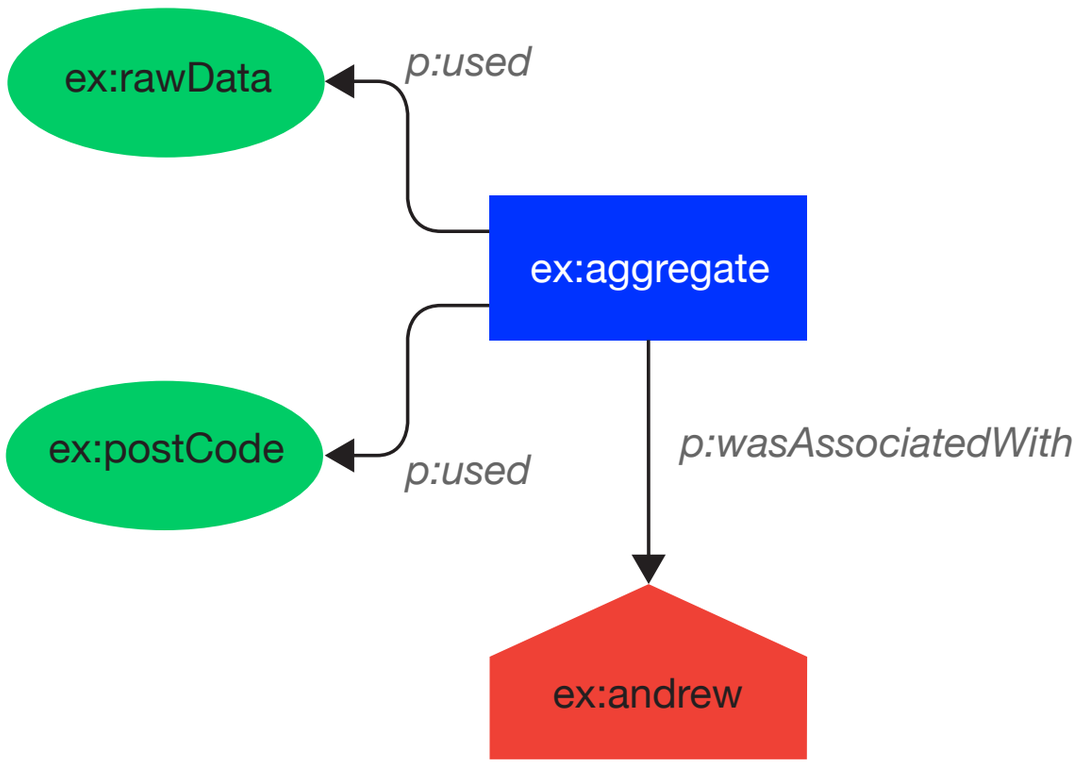
Entity

Agent

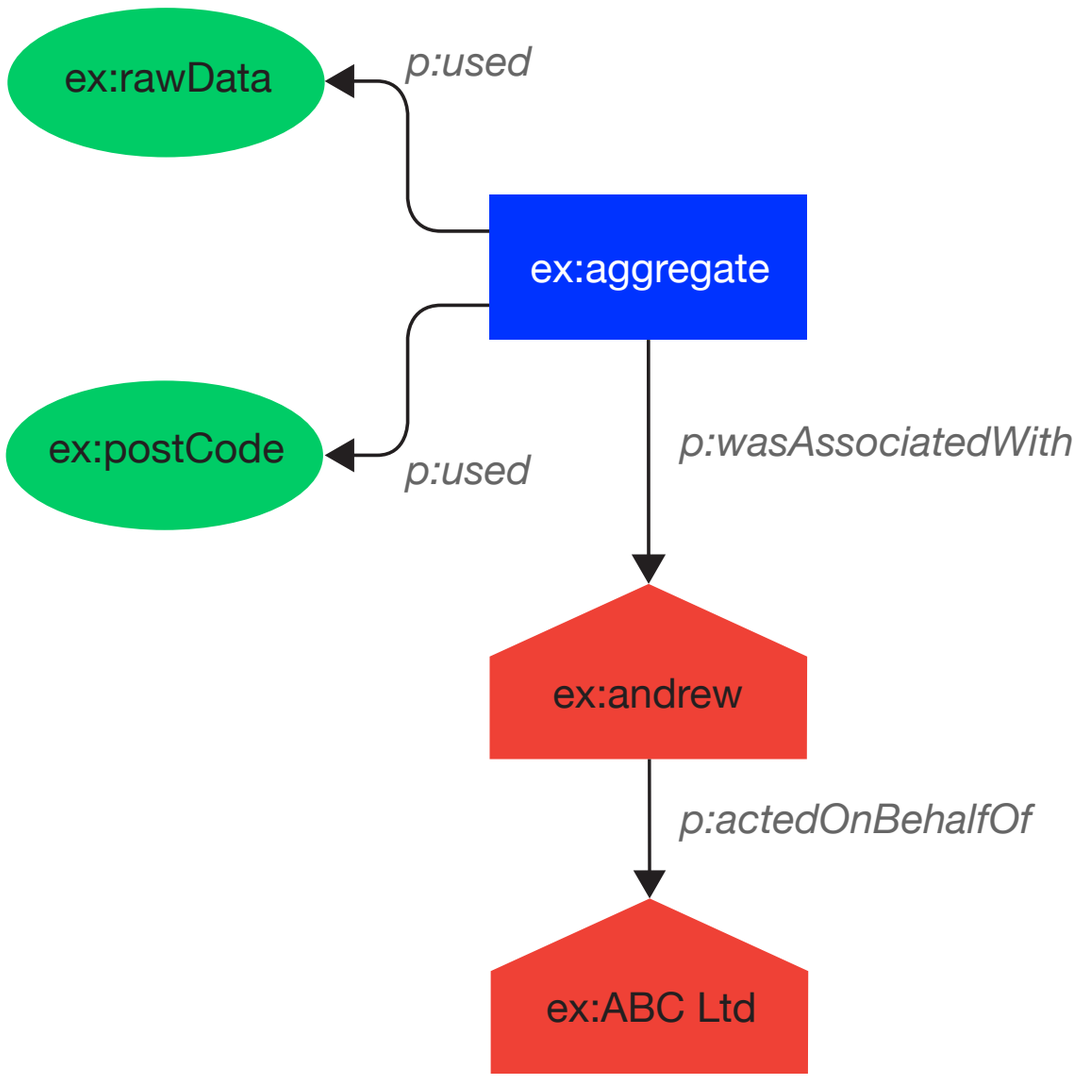
PROV-O EXAMPLE (6)



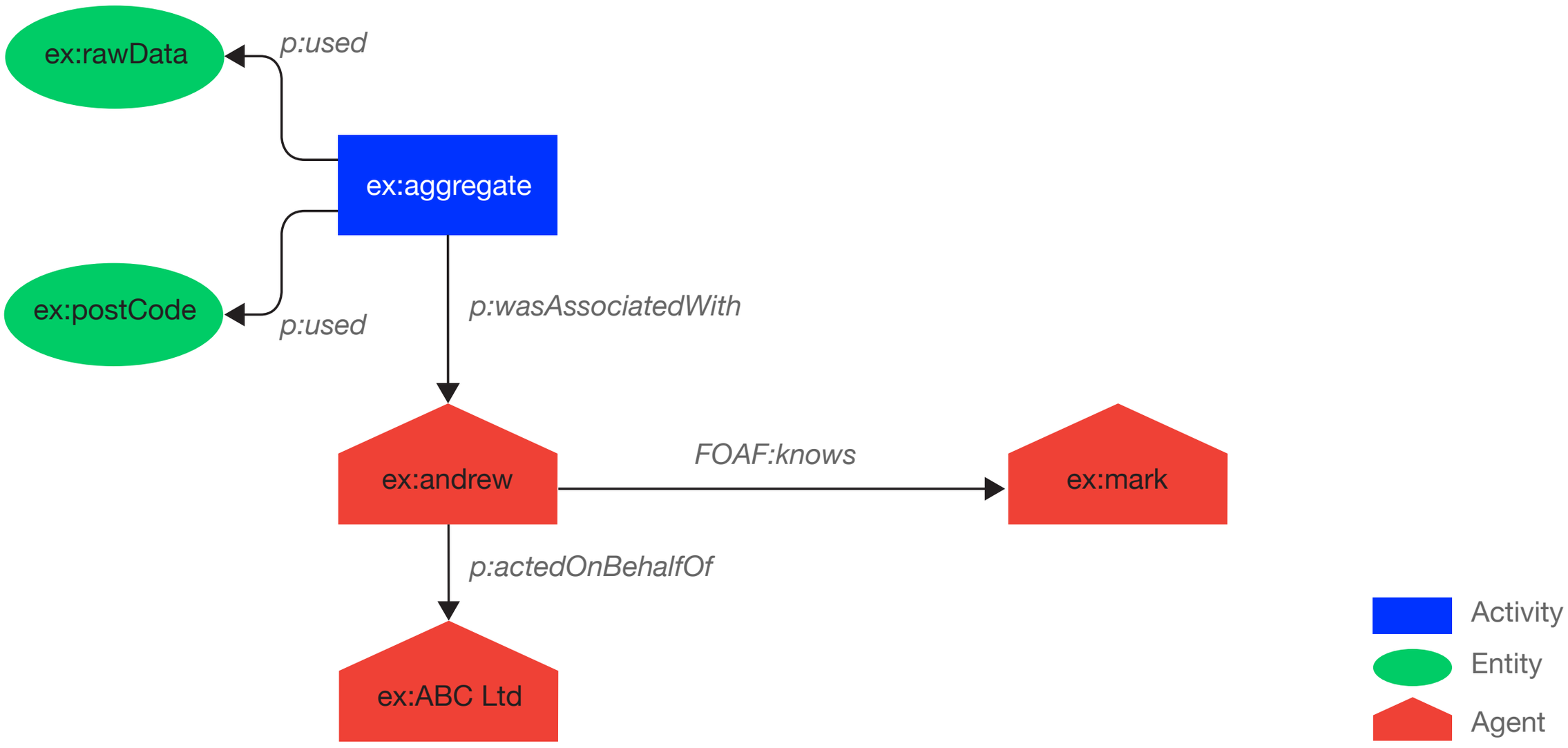
PROV-O EXAMPLE (7)



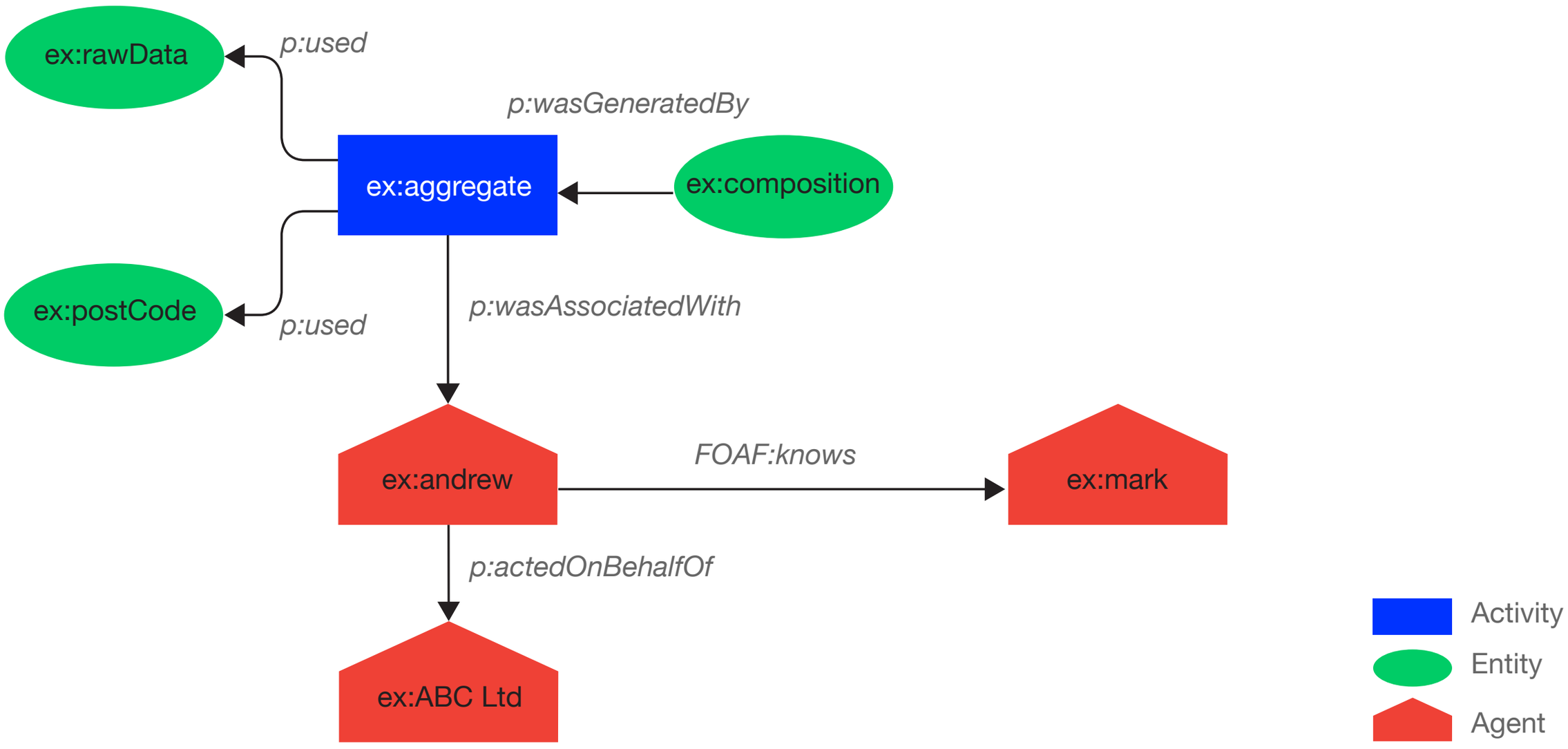
PROV-O EXAMPLE (8)



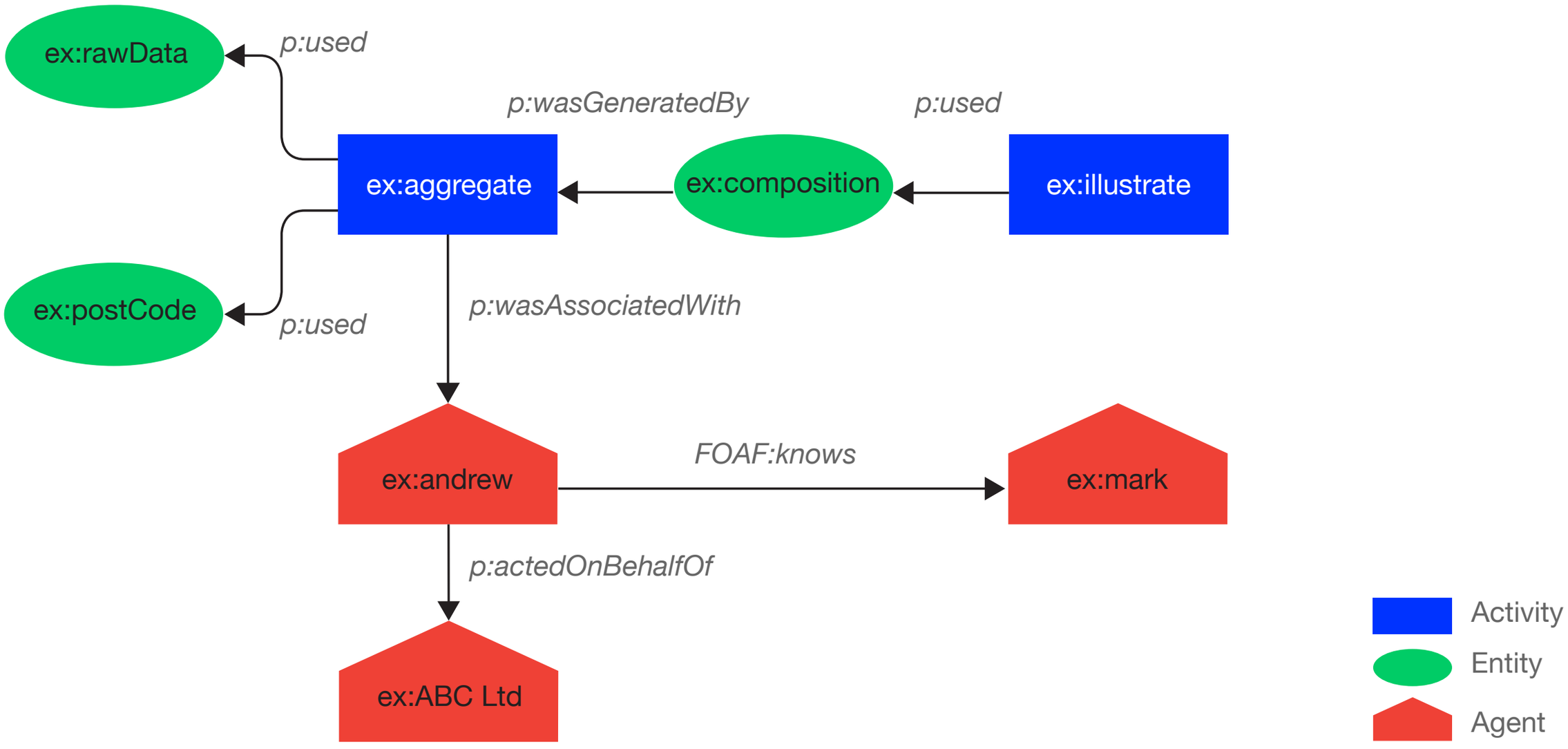
PROV-O EXAMPLE (9)



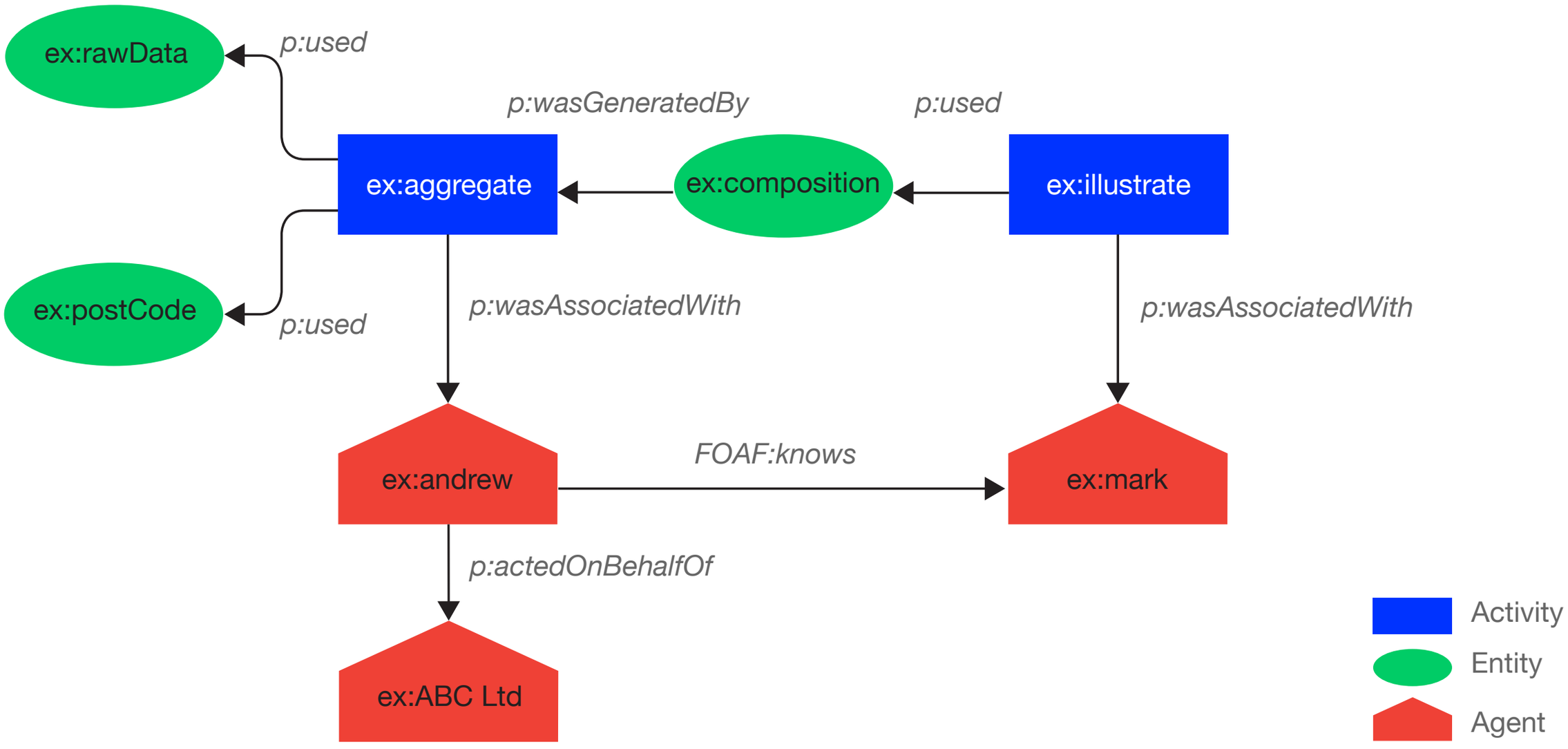
PROV-O EXAMPLE (10)



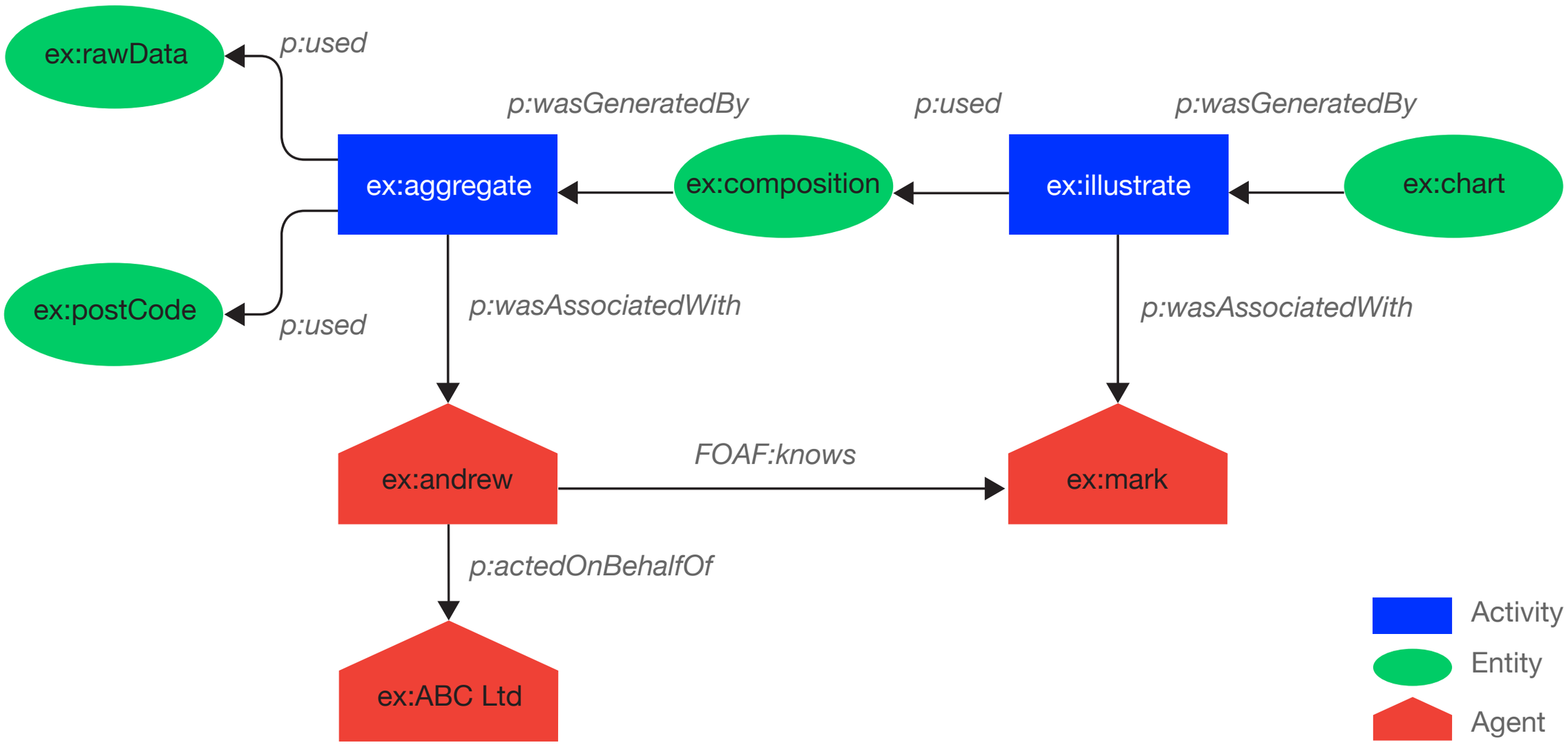
PROV-O EXAMPLE (11)



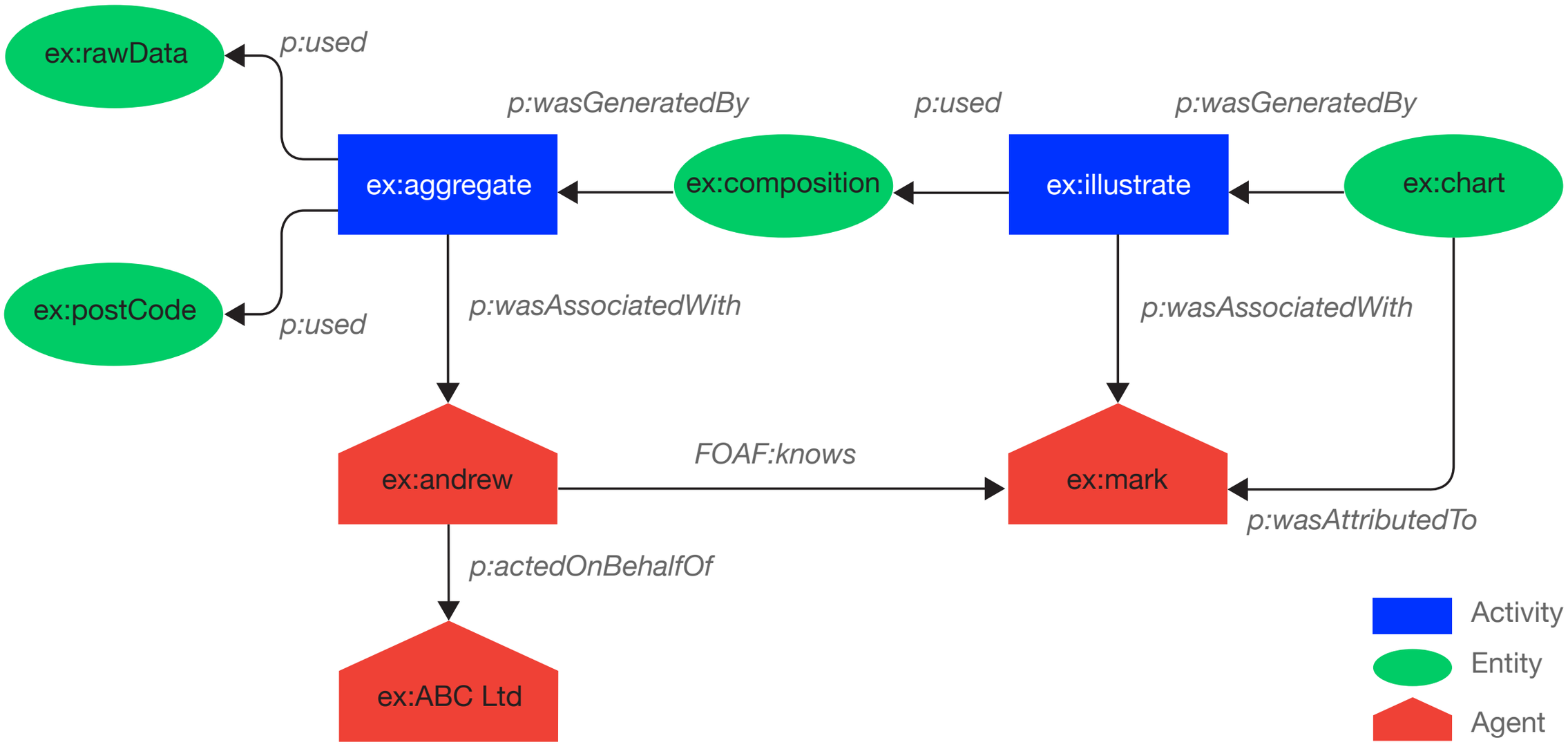
PROV-O EXAMPLE (12)



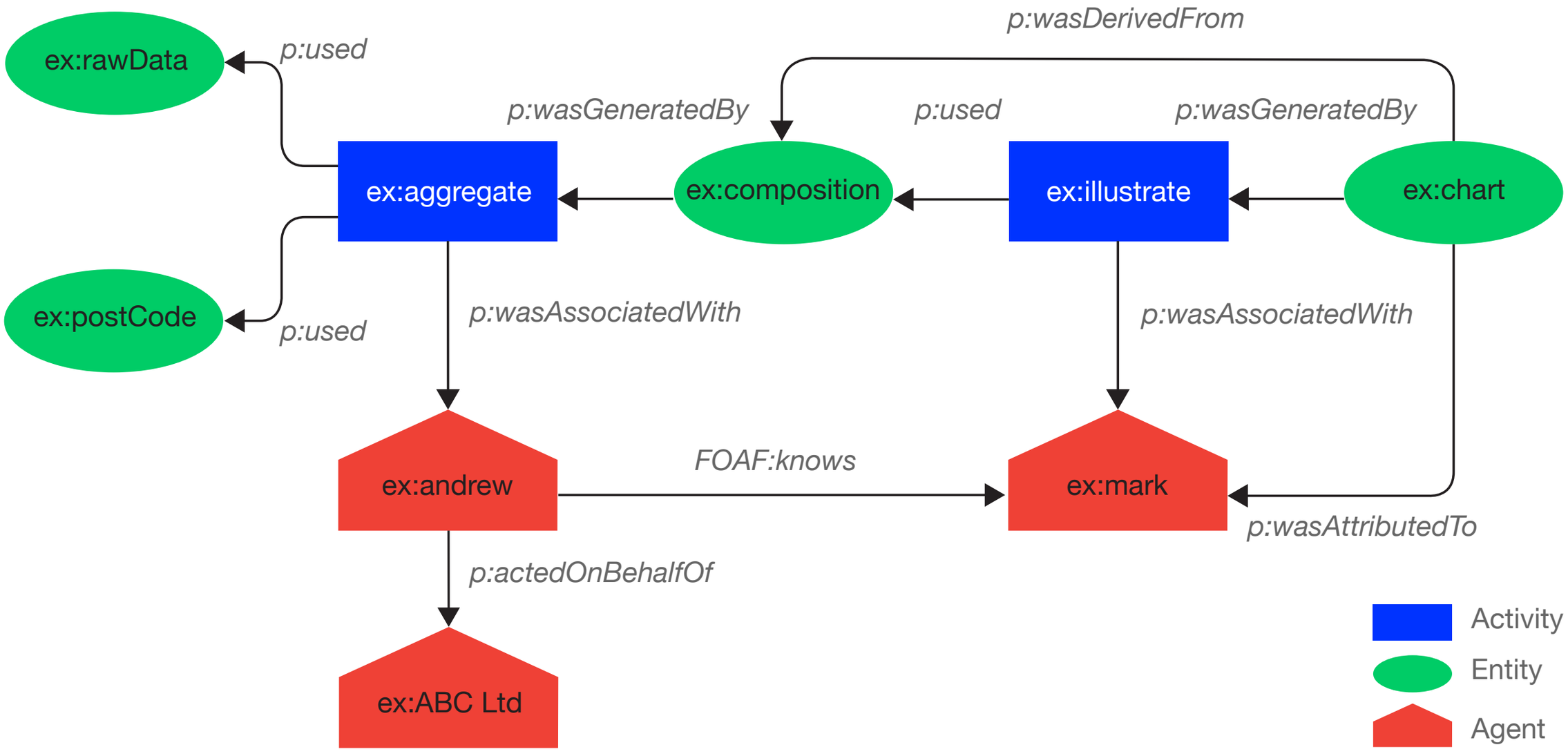
PROV-O EXAMPLE (13)



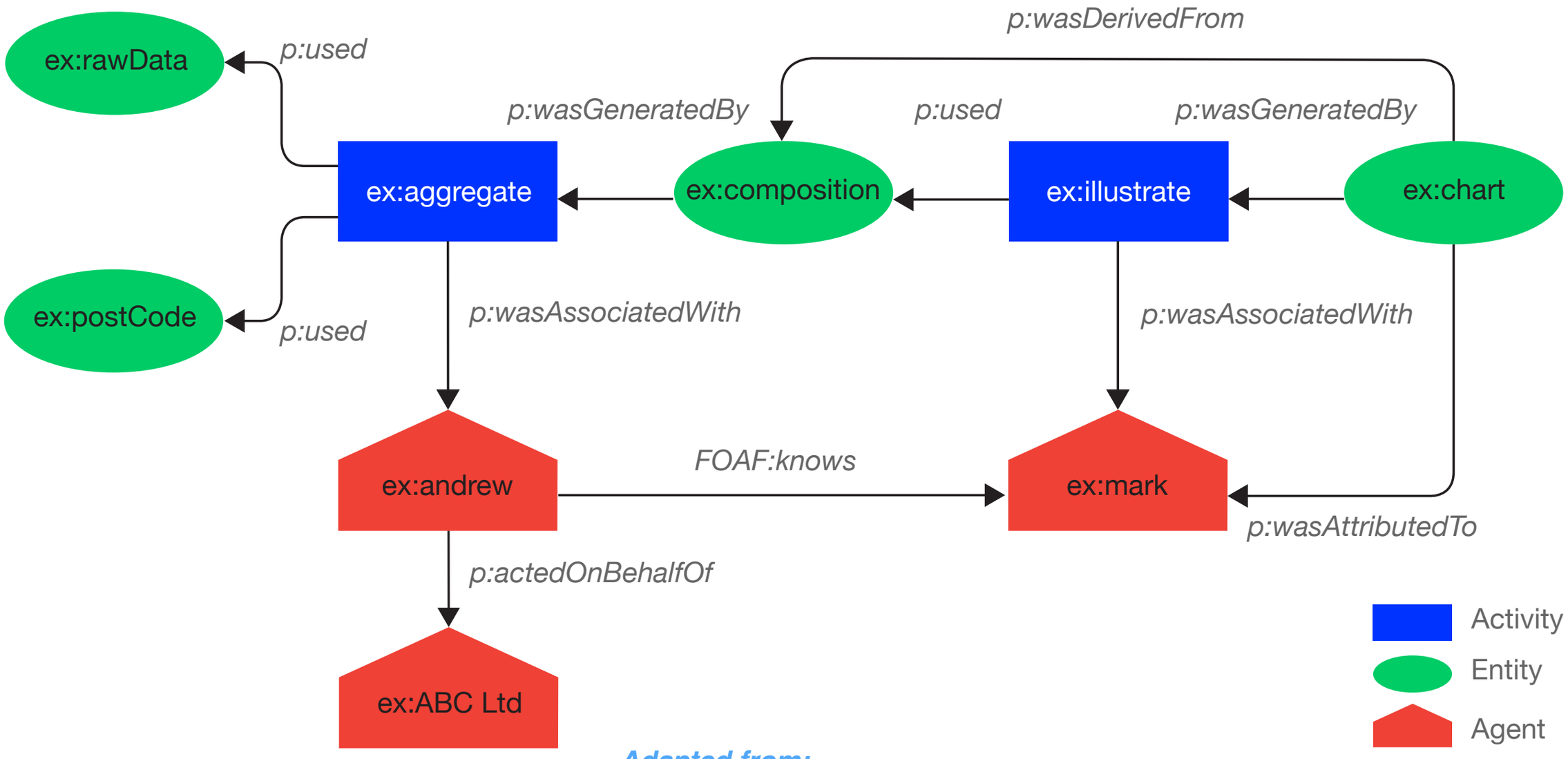
PROV-O EXAMPLE (14)



PROV-O EXAMPLE (15)



PROV-O EXAMPLE (16)



Adapted from:

http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/

COVERAGE OF PROV-O 'STARTING'

Properties

actedOnBehalfOf
wasAssociatedWith
wasAttributedTo
wasGeneratedBy
used
wasDerivedFrom

Not yet demonstrated

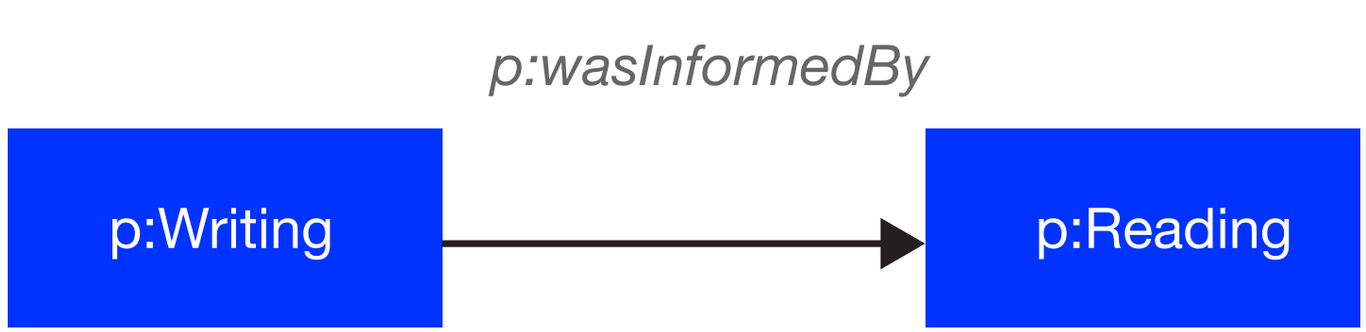
startedAtTime
endedAtTime
wasInformedBy

Classes

entity
activity
agent

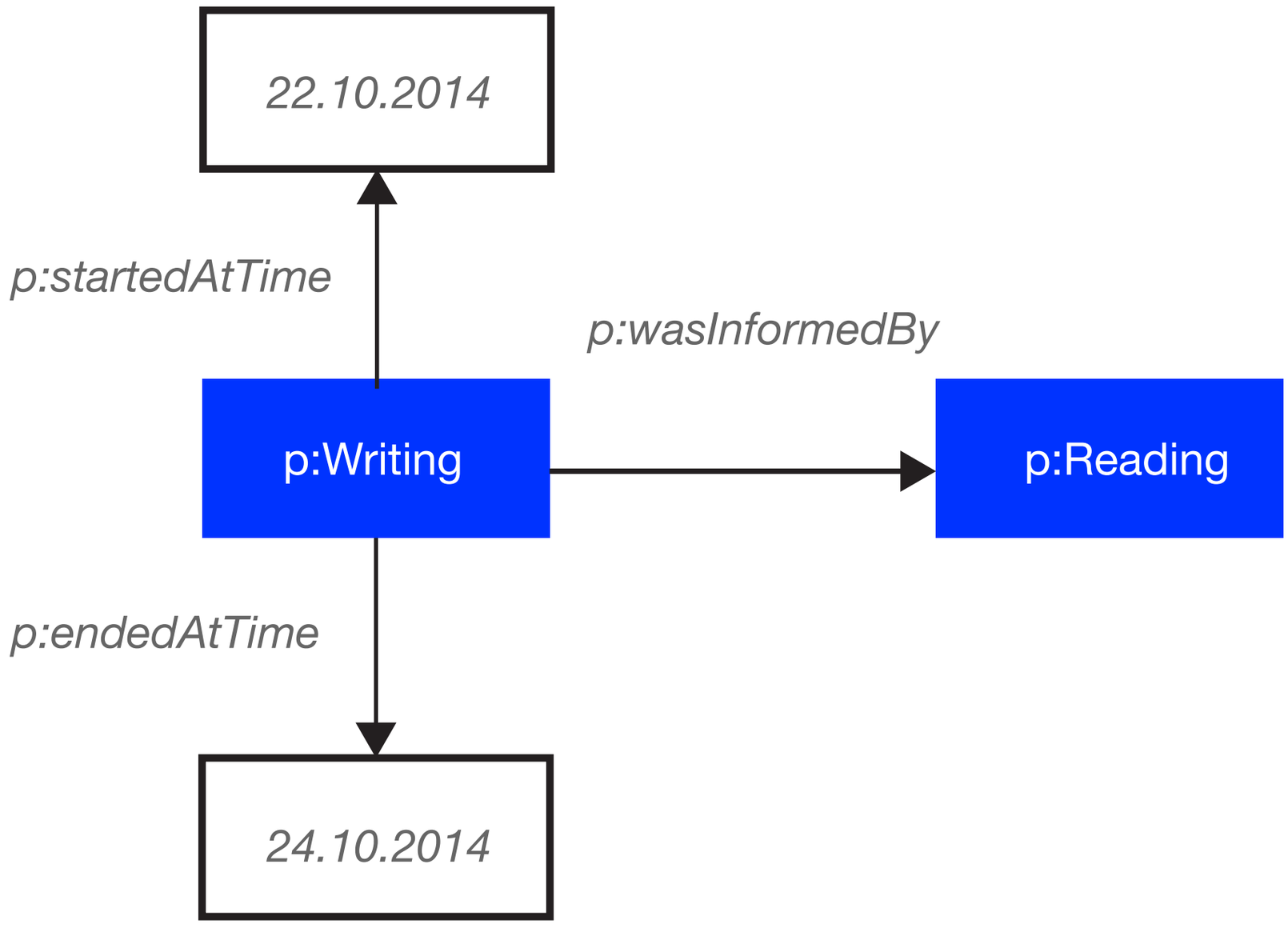
wasInformedBy

Connection between two activities



intermediate entity unspecified

TIME


22.10.2014
p:startedAtTime
p:wasInformedBy
p:Writing
p:Reading
p:endedAtTime
24.10.2014

BEYOND PROV-O ‘STARTING’

PROV-O 'Expanded'

Classes

prov:Collection prov:EmptyCollection prov:Bundle prov:Person prov:SoftwareAgent prov:Organization prov:Location

Properties

prov:alternateOf prov:specializationOf prov:generatedAtTime prov:hadPrimarySource prov:value prov:wasQuotedFrom prov:wasRevisionOf prov:invalidatedAtTime prov:wasInvalidatedBy prov:hadMember prov:wasStartedBy prov:wasEndedBy prov:invalidated prov:influenced prov:atLocation prov:generated

PROV-O 'Qualified'

Classes

prov:Influence prov:EntityInfluence prov:Usage prov:Start prov:End prov:Derivation prov:PrimarySource prov:Quotation prov:Revision prov:ActivityInfluence prov:Generation prov:Communication prov:Invalidation prov:AgentInfluence prov:Attribution prov:Association prov:Plan prov:Delegation prov:InstantaneousEvent prov:Role

Properties

prov:wasInfluencedBy prov:qualifiedInfluence prov:qualifiedGeneration prov:qualifiedDerivation prov:qualifiedPrimarySource prov:qualifiedQuotation prov:qualifiedRevision prov:qualifiedAttribution prov:qualifiedInvalidation prov:qualifiedStart prov:qualifiedUsage prov:qualifiedCommunication prov:qualifiedAssociation prov:qualifiedEnd prov:qualifiedDelegation prov:influencer prov:entity prov:hadUsage prov:hadGeneration prov:activity prov:agent prov:hadPlan prov:hadActivity prov:atTime prov:hadRole

END!