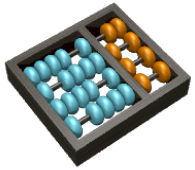
Handling Multiple Foci in Graph Databases



Institute of Computing

Jaudete Daltio
Claudia Bauzer Medeiros
(Advisor)

Outline

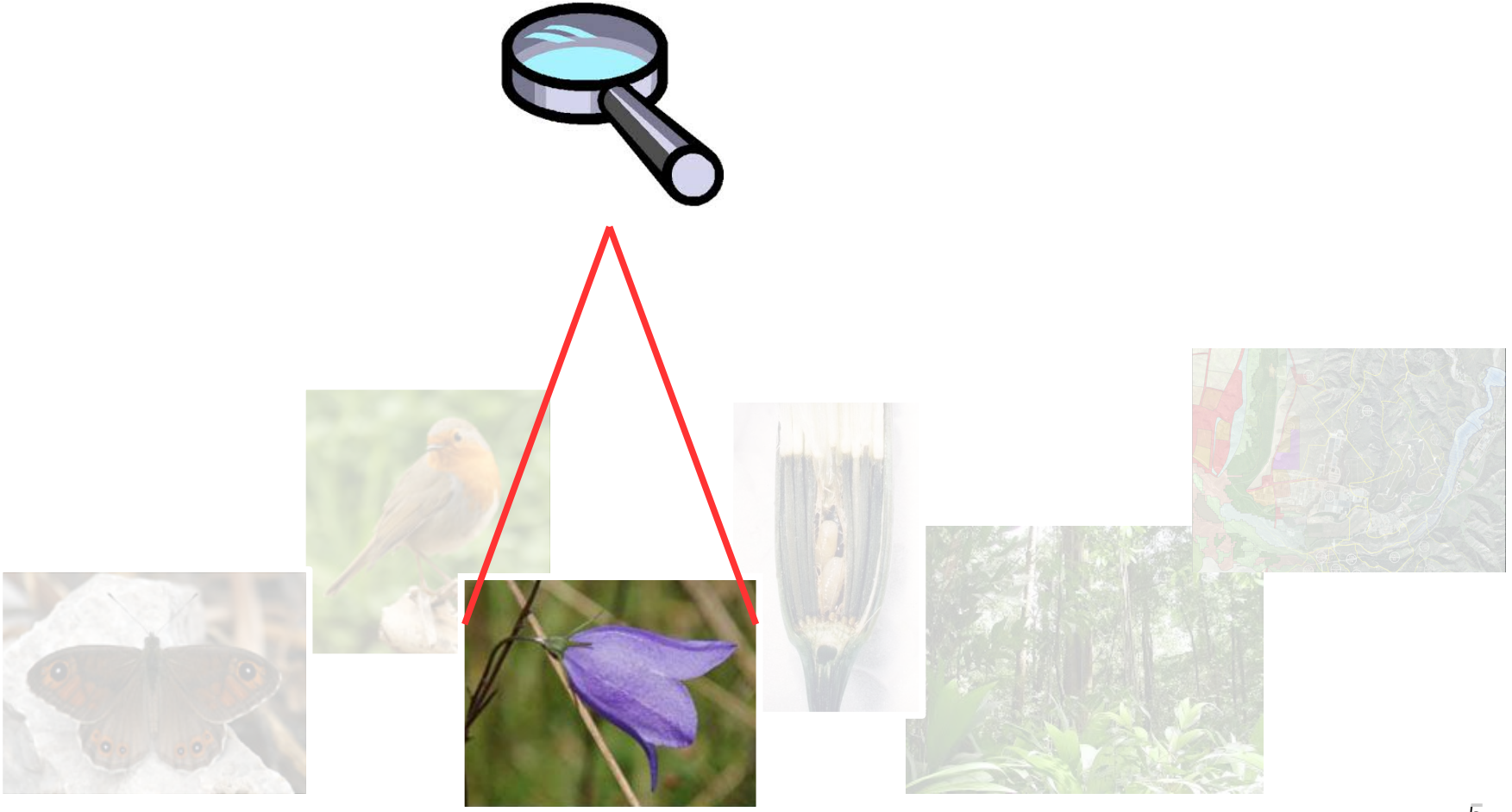
- Motivation
- Goal
- Proposal
- Running Example
- Ongoing Work

Biodiversity

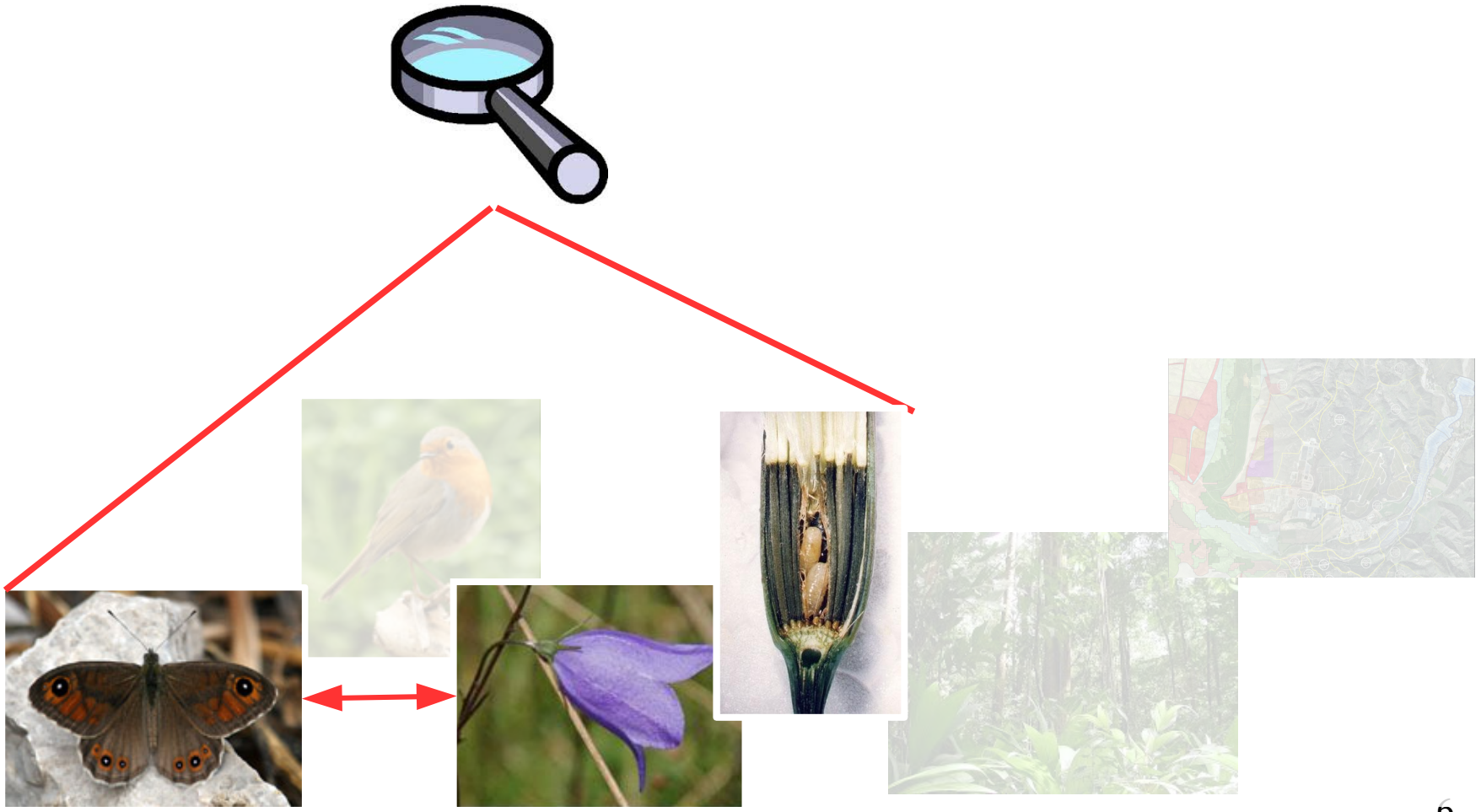
- Distributions and interactions between species, communities, ecosystems and biomes
- Identification of possible scenarios from the local to the global level (scale)
- Several experiments are needed to check scientific hypothesis (multiple perspectives)

Biodiversity

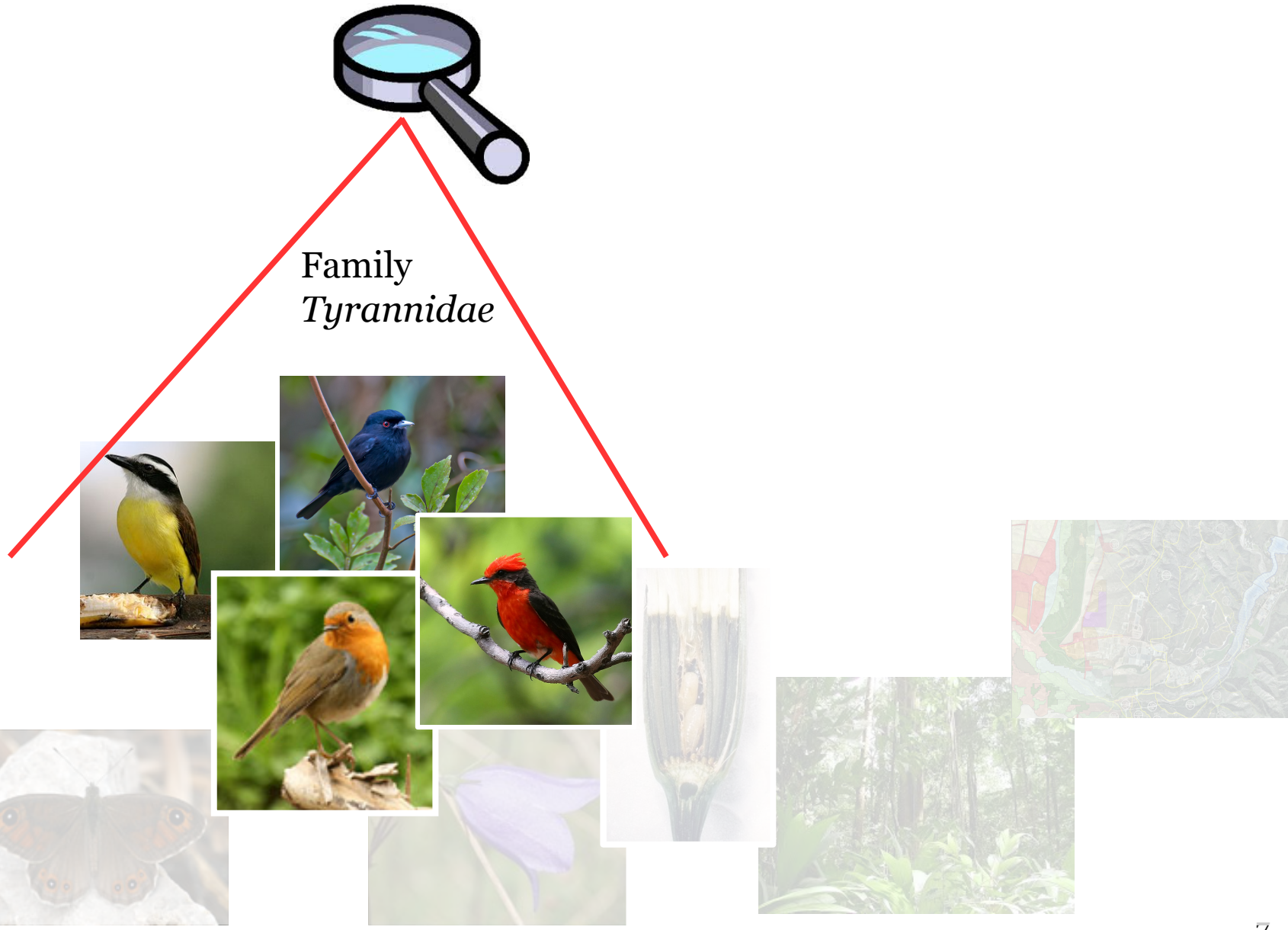
Biodiversity



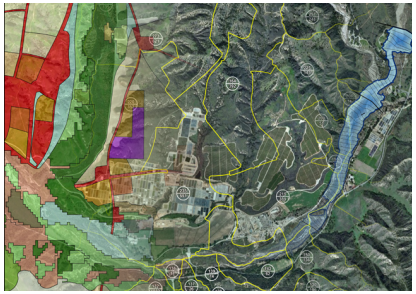
Biodiversity



Biodiversity



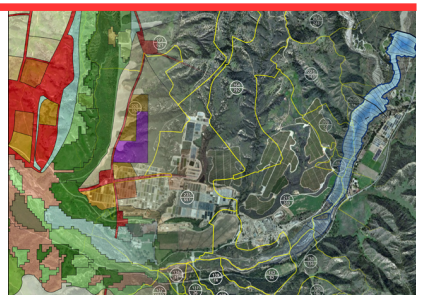
Biodiversity



Location

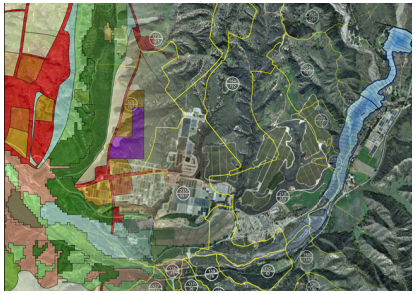
Biodiversity

Region



Biodiversity

Biome Temperate Deciduous Forest



Biodiversity

- Distributions and interactions between species, communities, ecosystems and biomes
- Identification of possible scenarios from the local to the global level (scale)
- Several experiments are needed to check scientific hypothesis (multiple perspectives)

Goal

Build and explore arbitrary foci in scientific databases

- Highly connected
- Different scales and shapes

Goal

Build and explore arbitrary foci in scientific databases

- Highly connected
- Different scales and shapes

Focus

- Perspective of study of a given problem [Santache et al. 2012]
- Restricts data and creates data subsets of interest
  - one specific scale
  - one specific representation
  - put together objects from distinct scales
- Analyze data under different models
- Process data using focus-specific algorithms

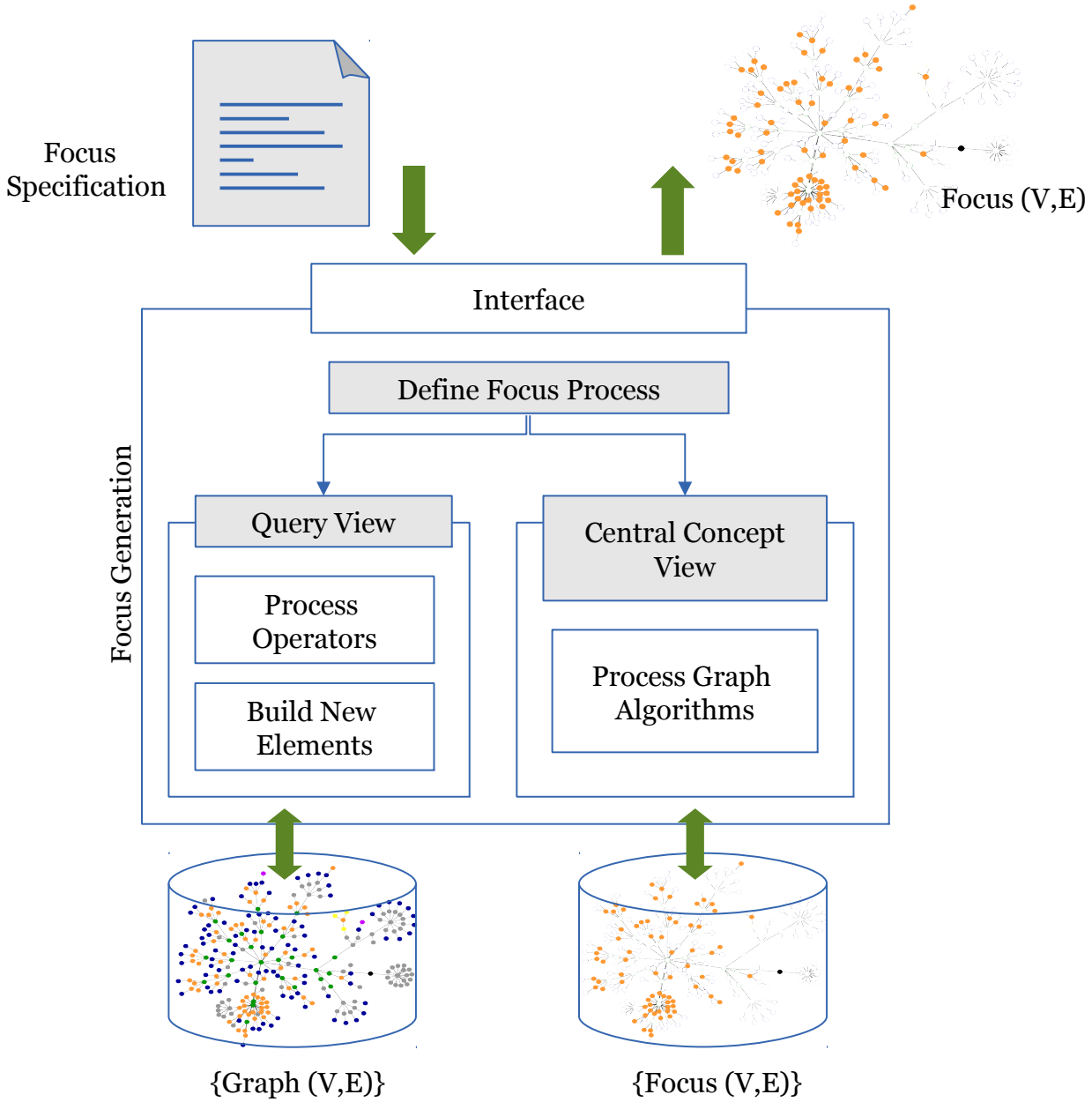
Goal

Build and explore arbitrary foci in scientific databases

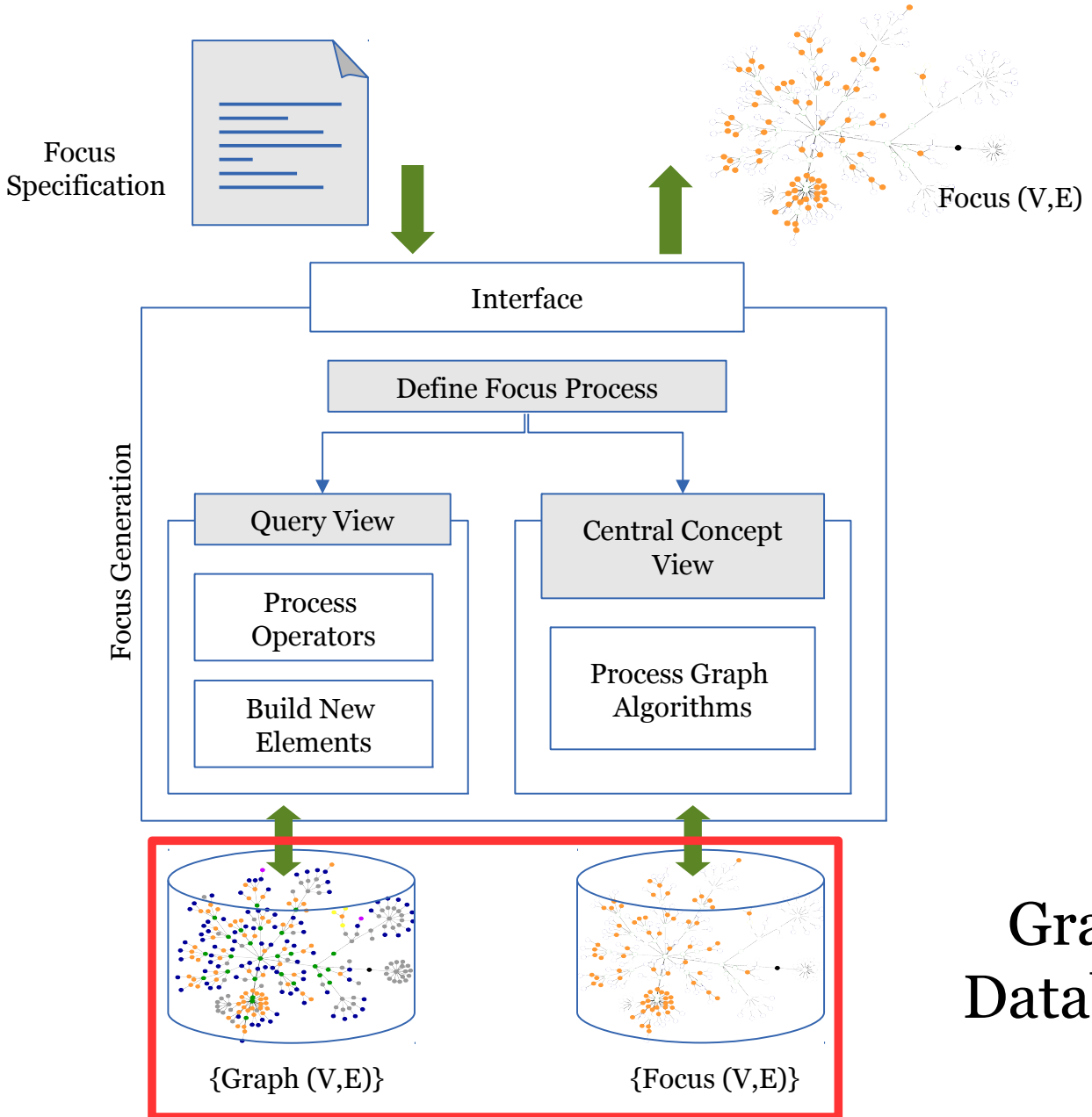
- Highly connected
- Different scales and shapes

Graph Databases + Views

Proposal

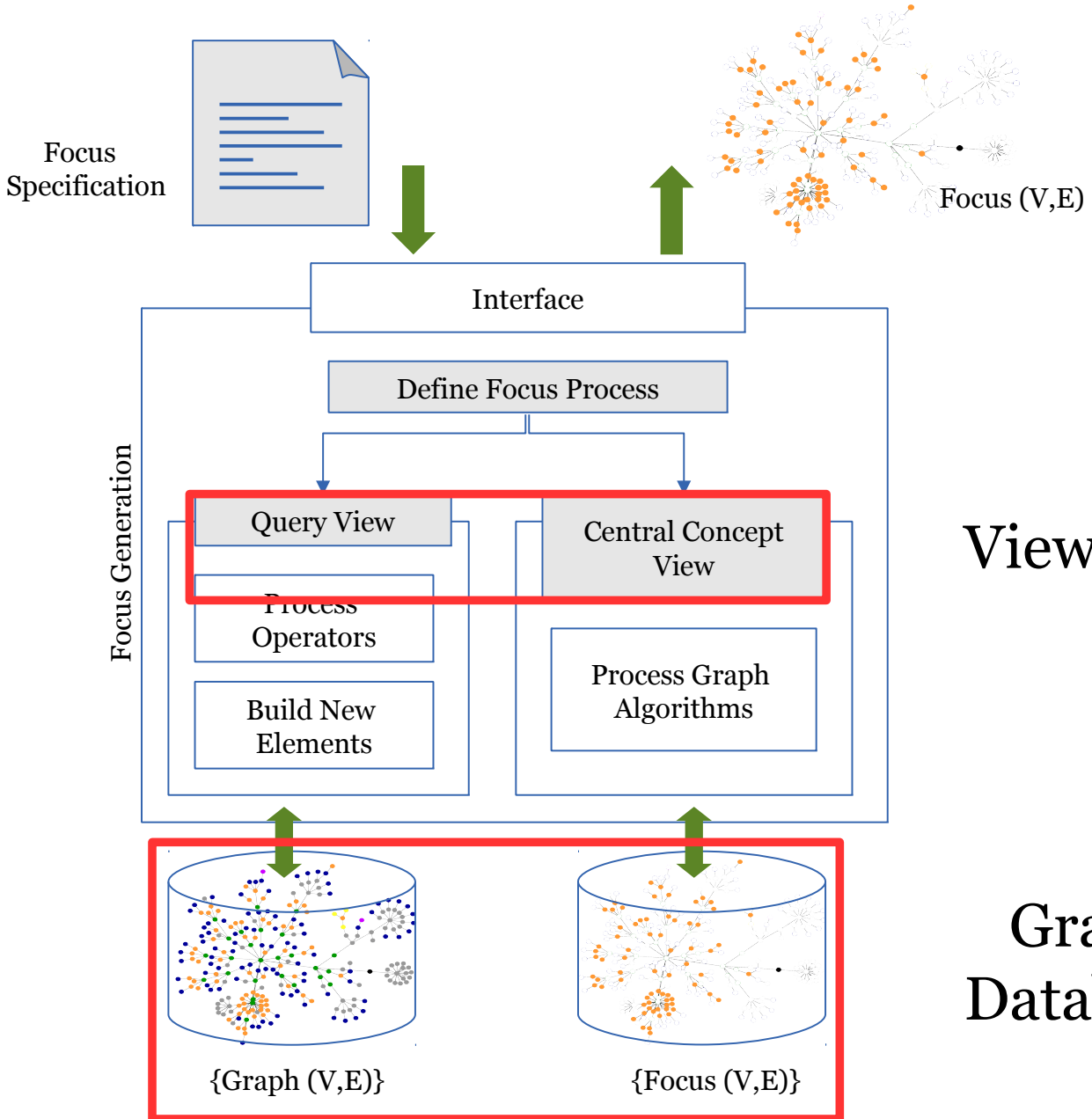

Focus Specification
Focus (V,E)
Interface
Define Focus Process
Focus Generation
Query View
Process Operators
Build New Elements
Central Concept View
Process Graph Algorithms
{Graph (V,E)}
{Focus (V,E)}

Proposal



Graph Databases

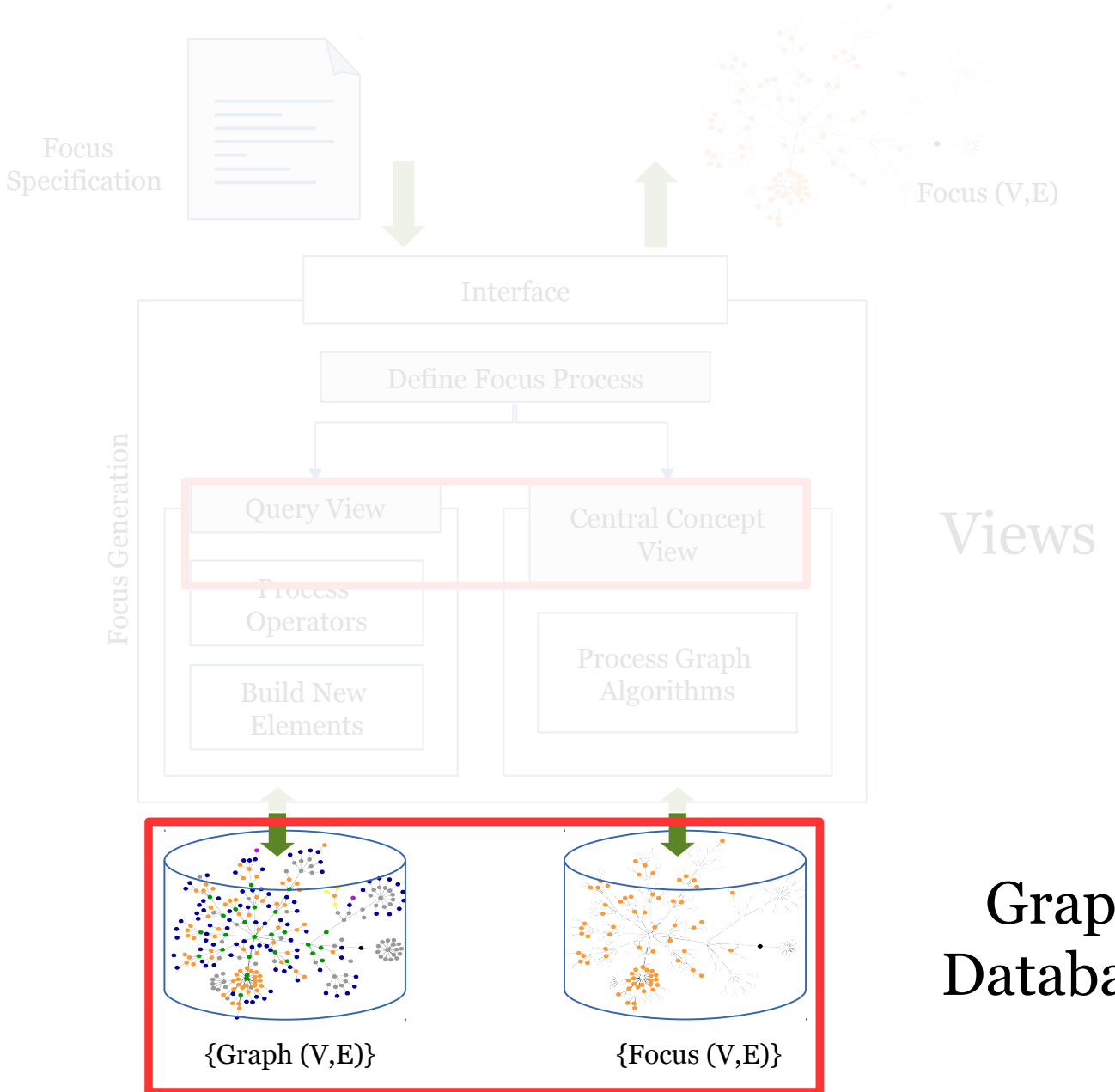
Proposal



Views

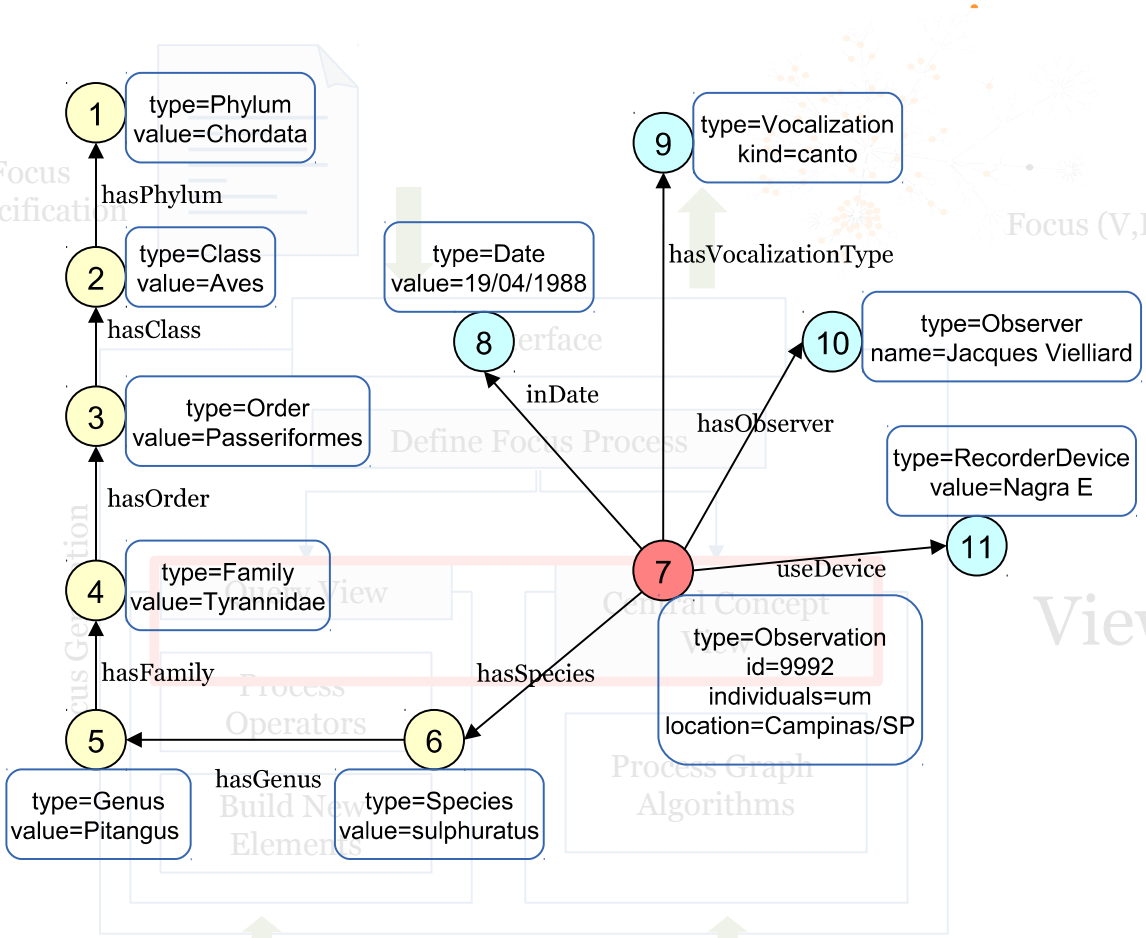
Graph Databases

Proposal



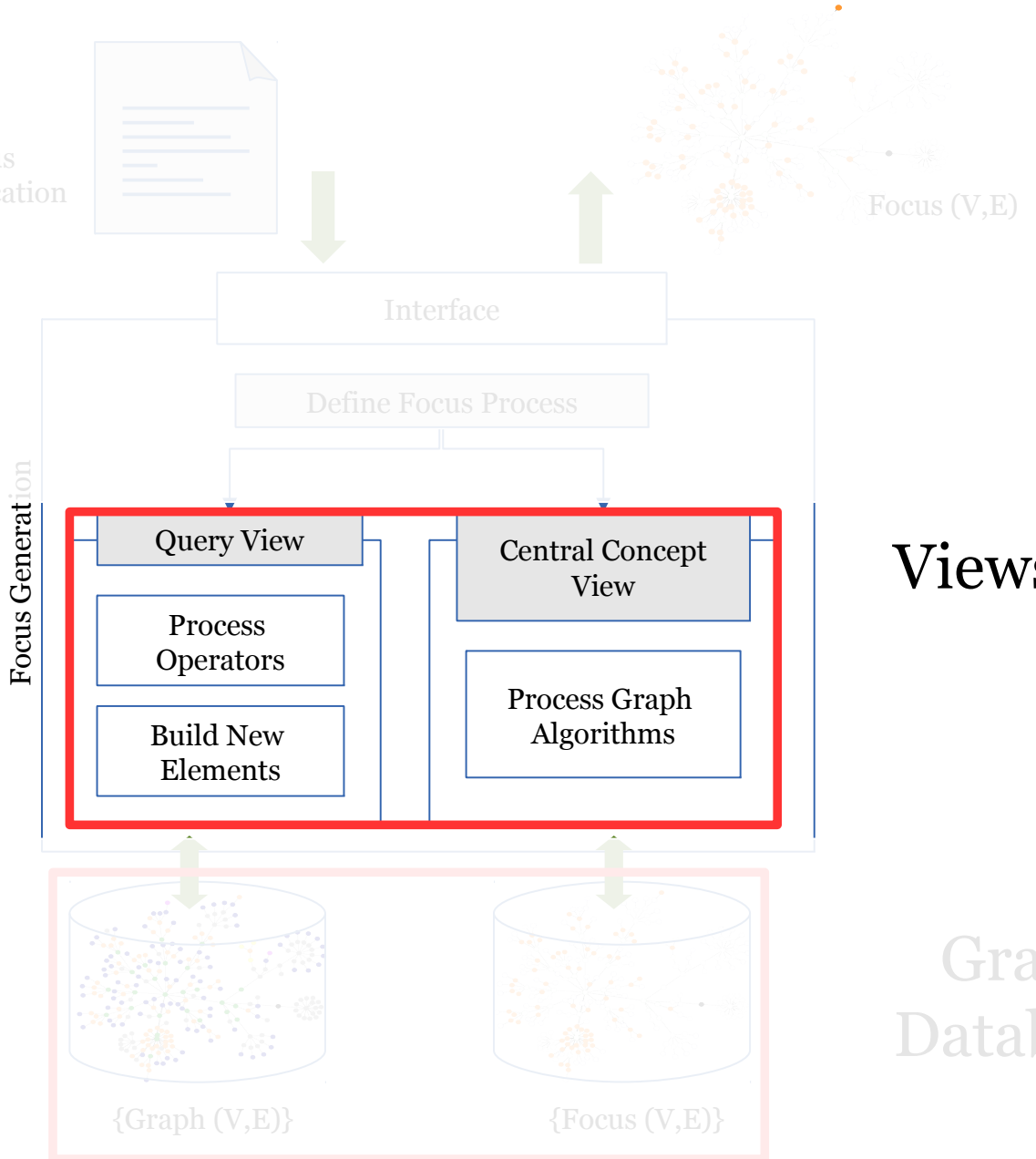
Graph Databases

Proposal



Property Graph

Proposal

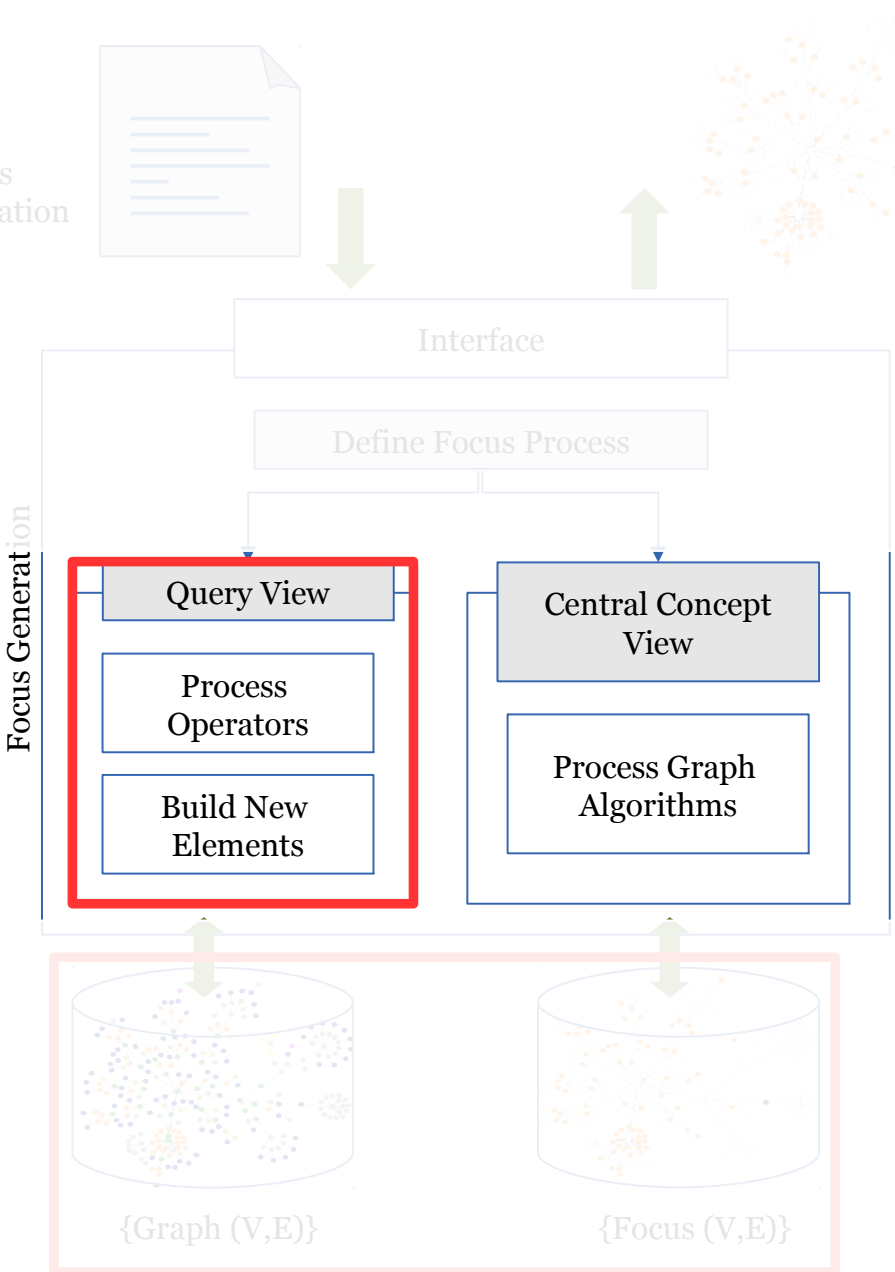


Views

Graph Databases

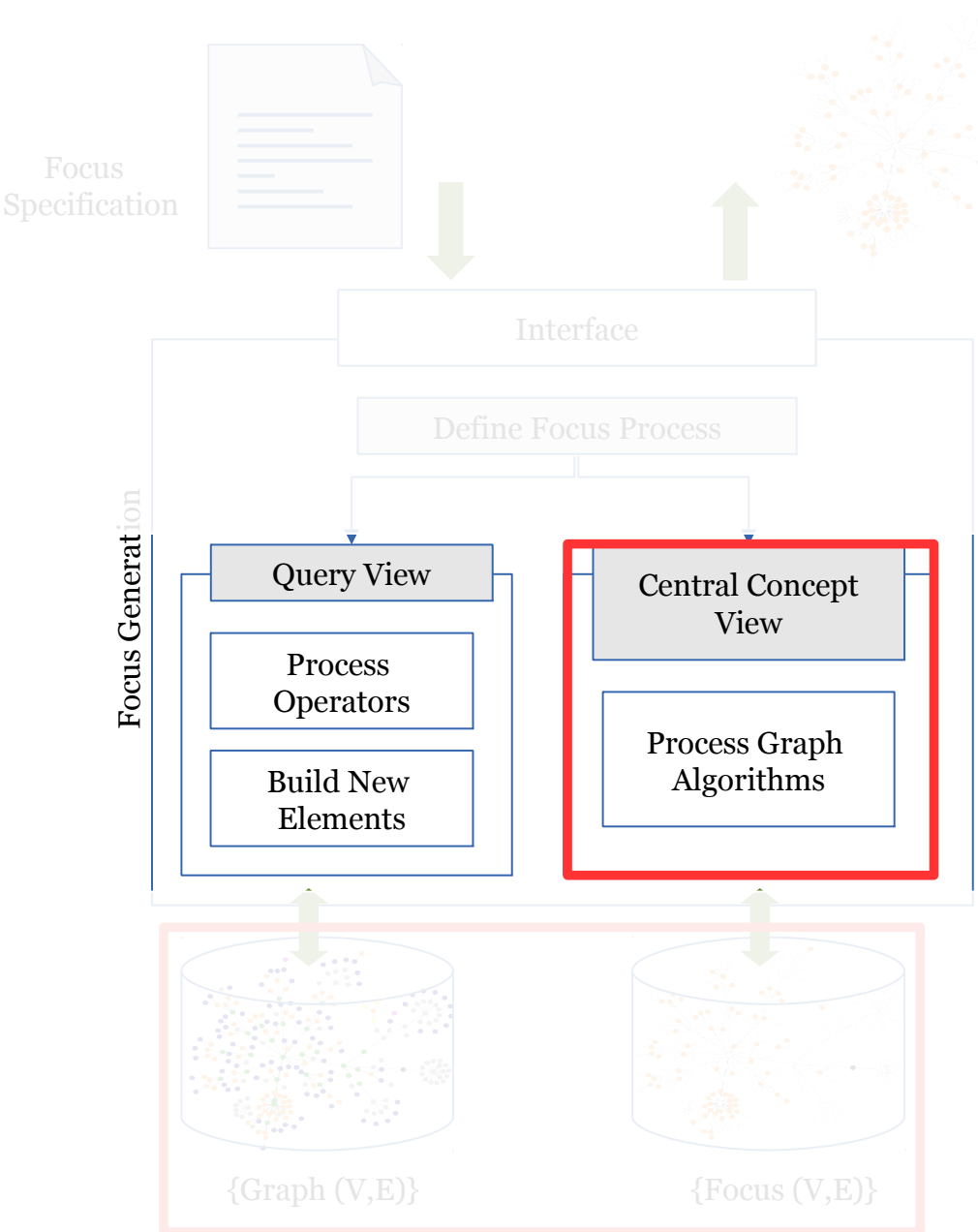
Proposal

Relational Databases (selection, projection, join, aggregate Functions)
+
Build New Elements



Graph Databases

Proposal



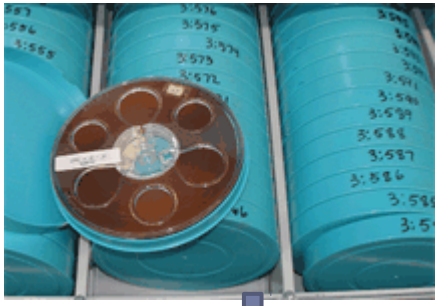
Ontology
(central concept)
+
Graph
Algorithms

Graph Databases

Running Example

- Observations: recordings of animal sounds
- Over 30 thousand recordings

Running Example - FNJV

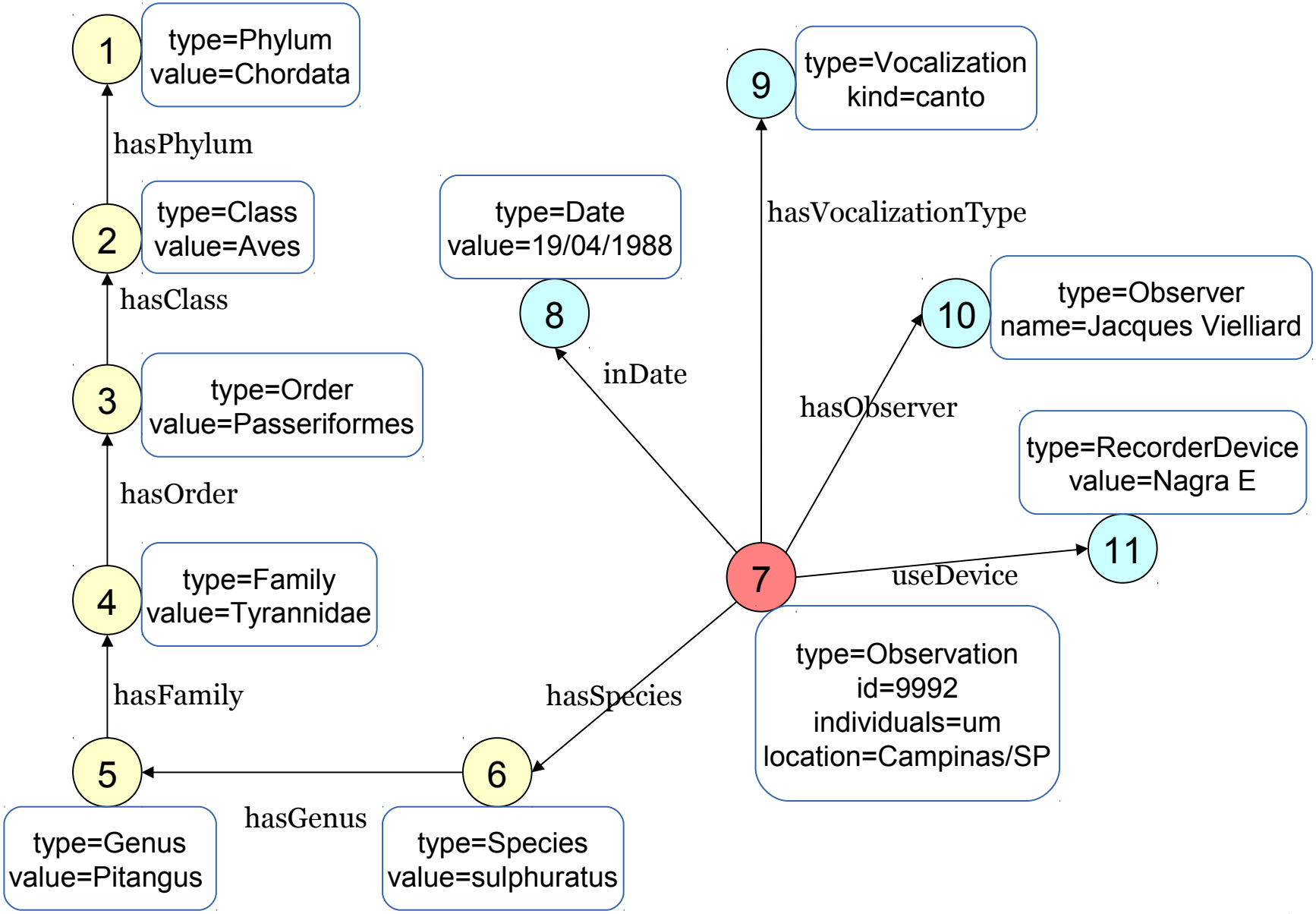


- Scientific Taxonomy
- Collect time/date/location
- Recordist
- Air Temperature
- Recording device model

54 metadata attributes

E	F	G	H	I
Filo	Classe	Ordem	Família	Gênero
Chordata	Aves	Sphenisciformes	Spheniscidae	Spheniscus
Chordata	Aves	Struthioniformes	Rheidae	Rhea
Chordata	Aves	Struthioniformes	Rheidae	Rhea
Chordata	Aves	Struthioniformes	Rheidae	Rhea
Chordata	Aves	Struthioniformes	Rheidae	Rhea
Chordata	Aves	Tinamiformes	Tinamidae	Tinamus
Chordata	Aves	Tinamiformes	Tinamidae	Tinamus

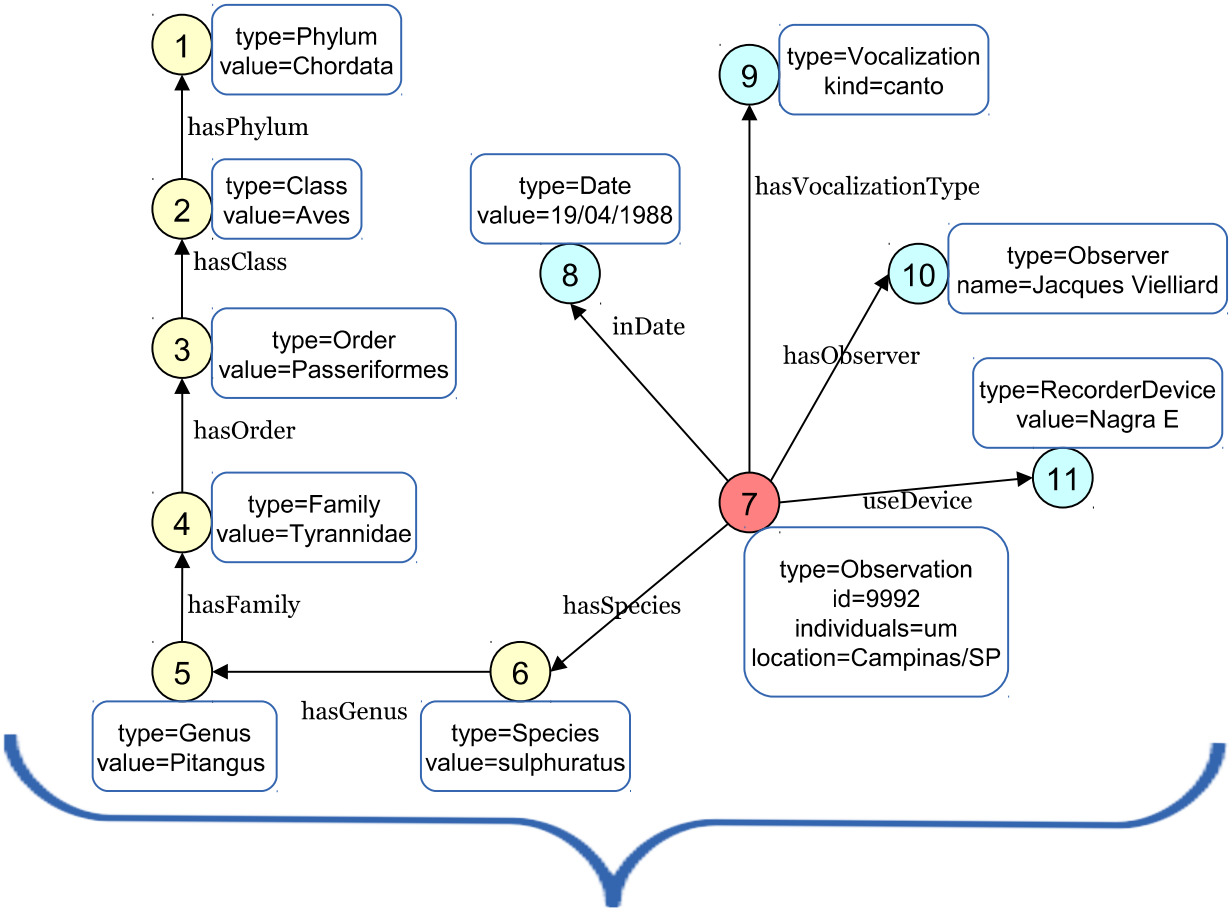
Running Example - FNJV



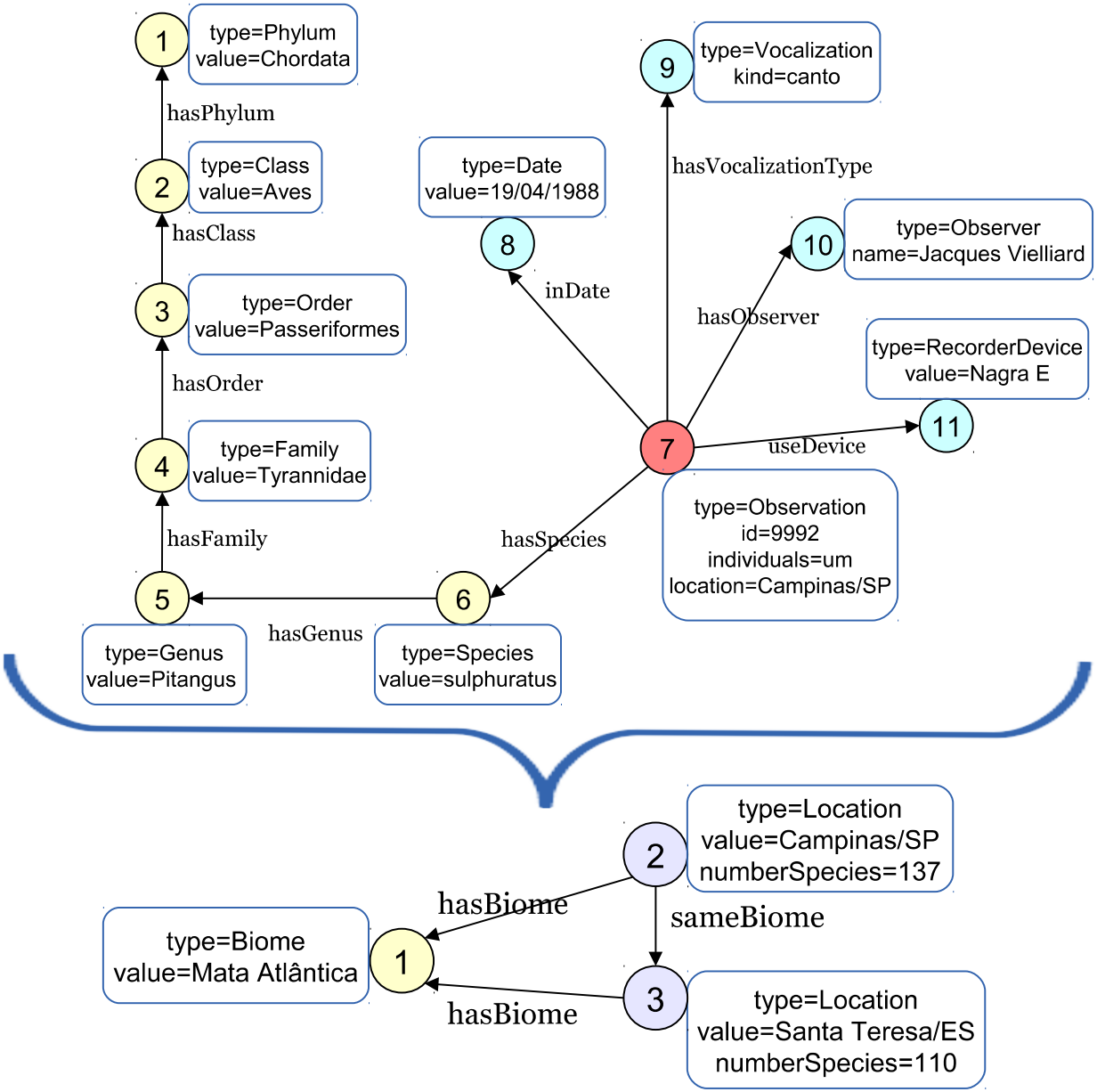
Focus 1

- *Set of all locations in which observations were made, summarizing the number of distinct species at each location, and connect the locations that belong to the same biome*
  - Location of observations
  - Number of Species
  - Biome
  - Same Biome

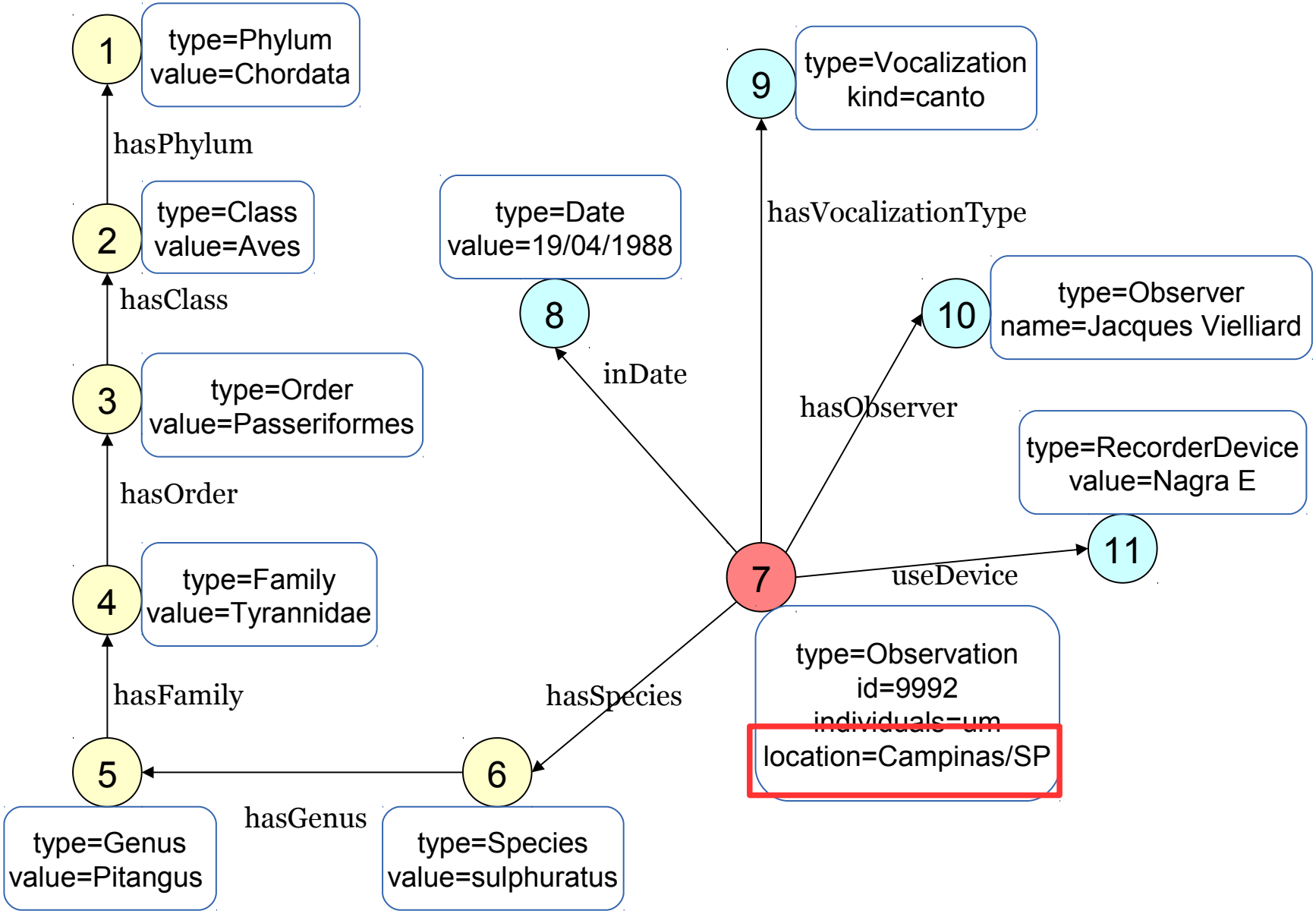
Focus 1


1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
type=Observation
id=9992
individuals=um
location=Campinas/SP
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice

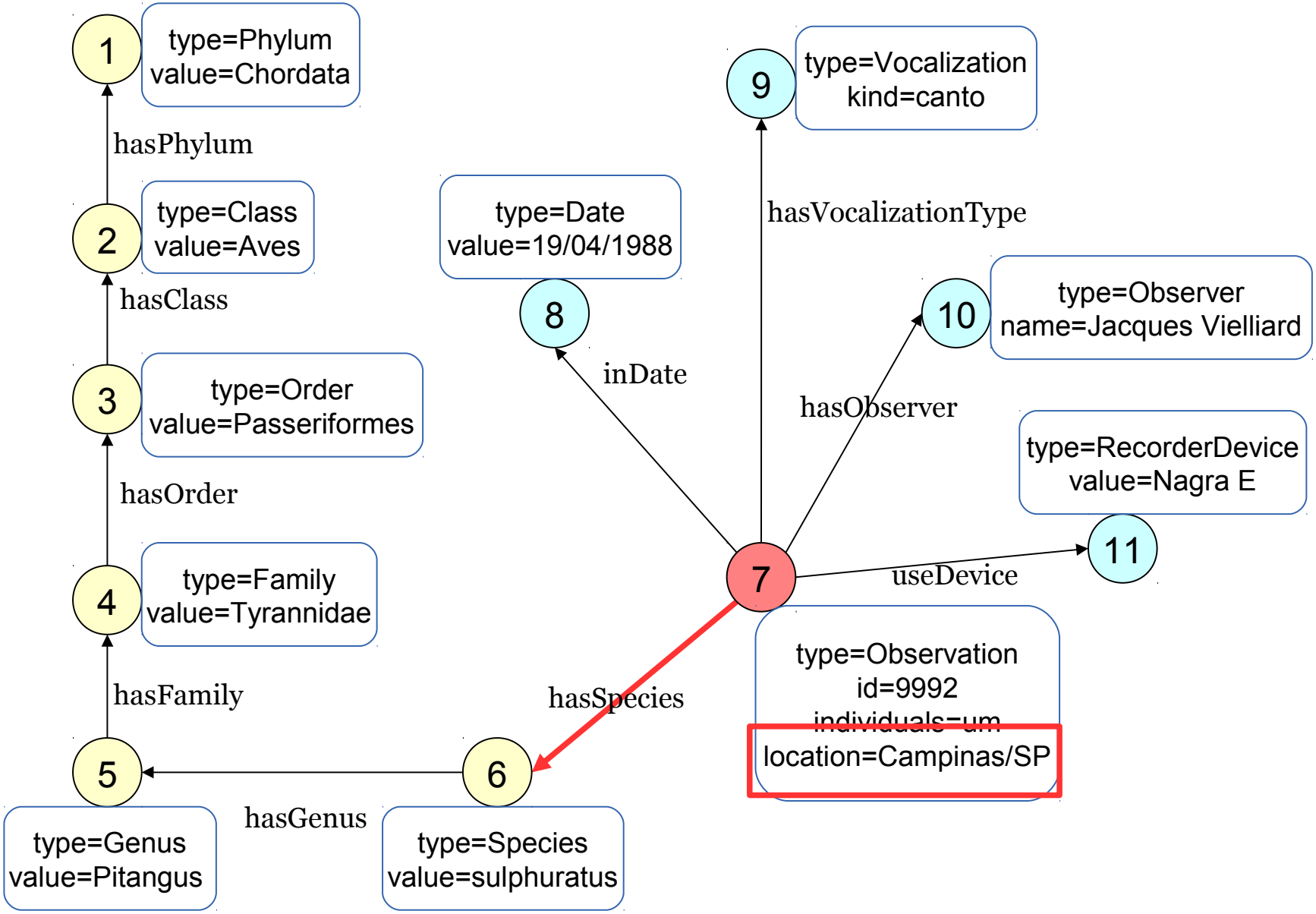
Focus 1


1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
type=Observation
id=9992
individuals=um
location=Campinas/SP
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice
1
type=Biome
value=Mata Atlântica
2
type=Location
value=Campinas/SP
numberSpecies=137
3
type=Location
value=Santa Teresa/ES
numberSpecies=110
hasBiome
hasBiome
sameBiome

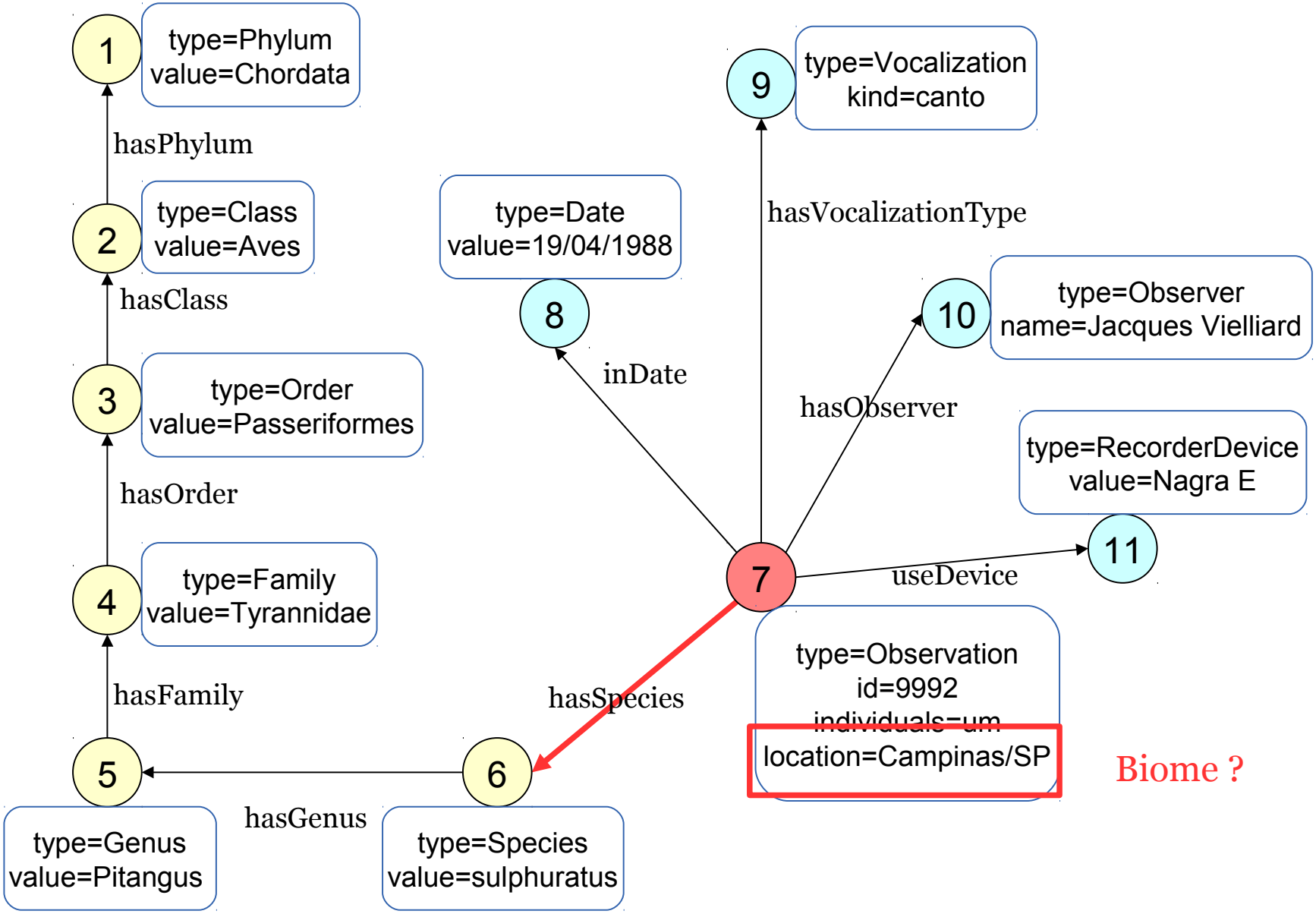
Observations Graph Database



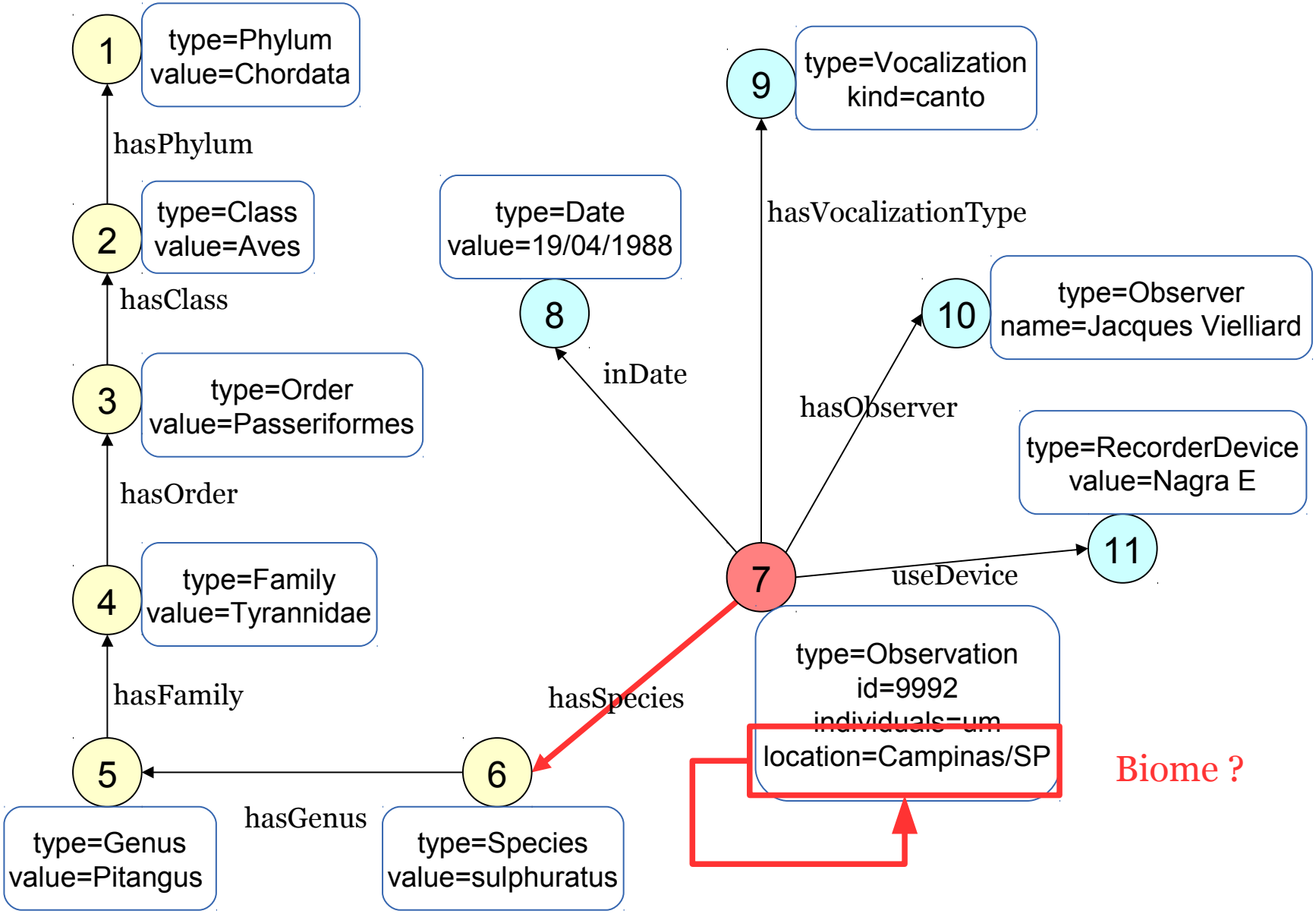
Observations Graph Database



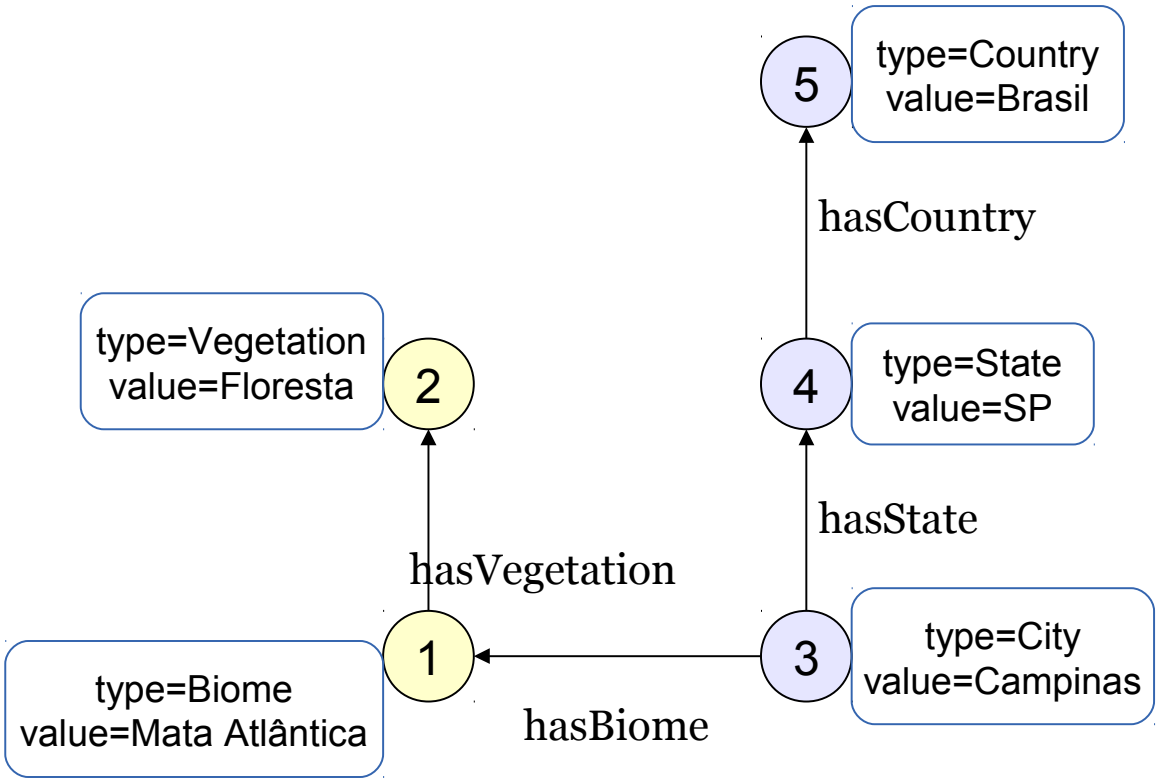
Observations Graph Database



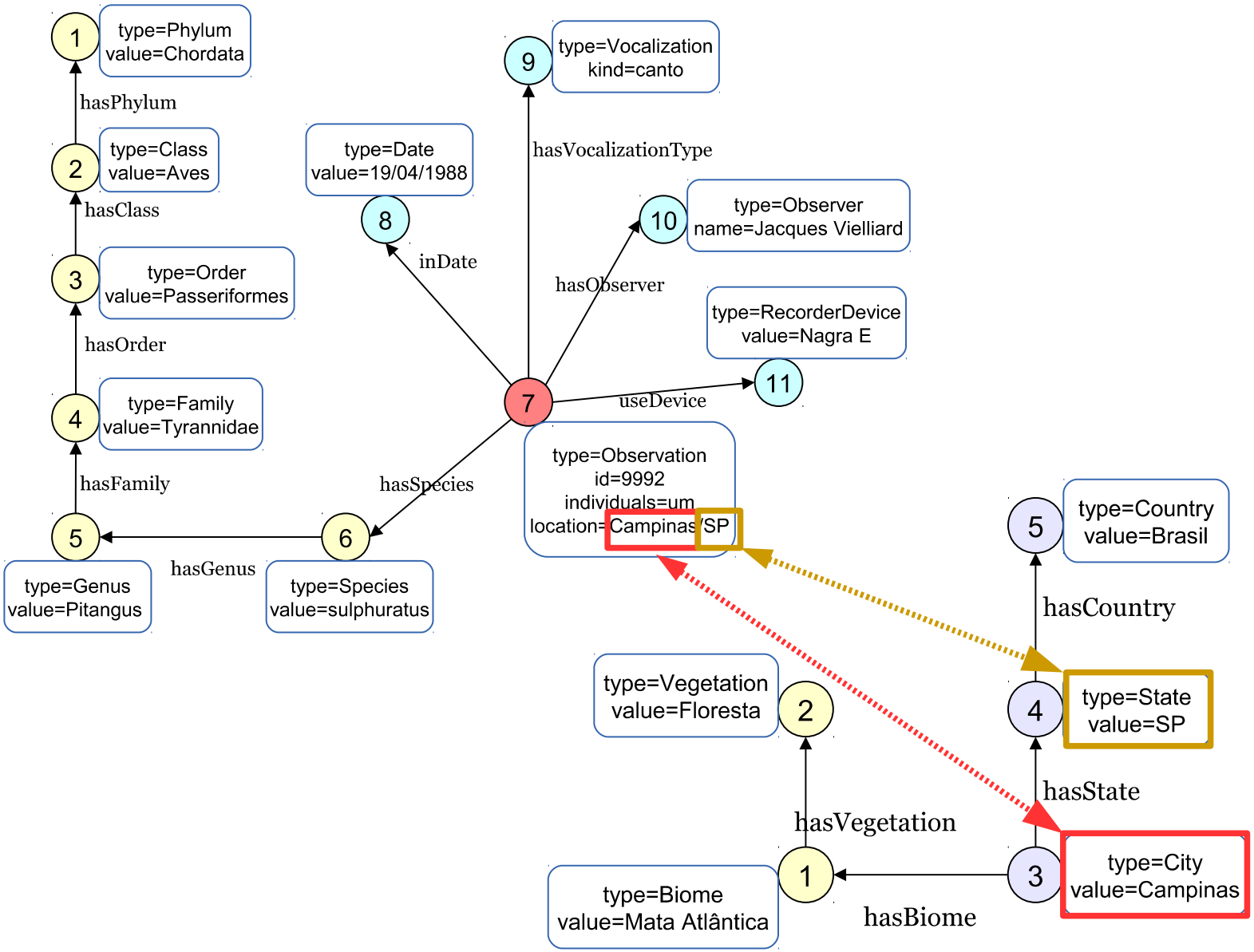
Observations Graph Database



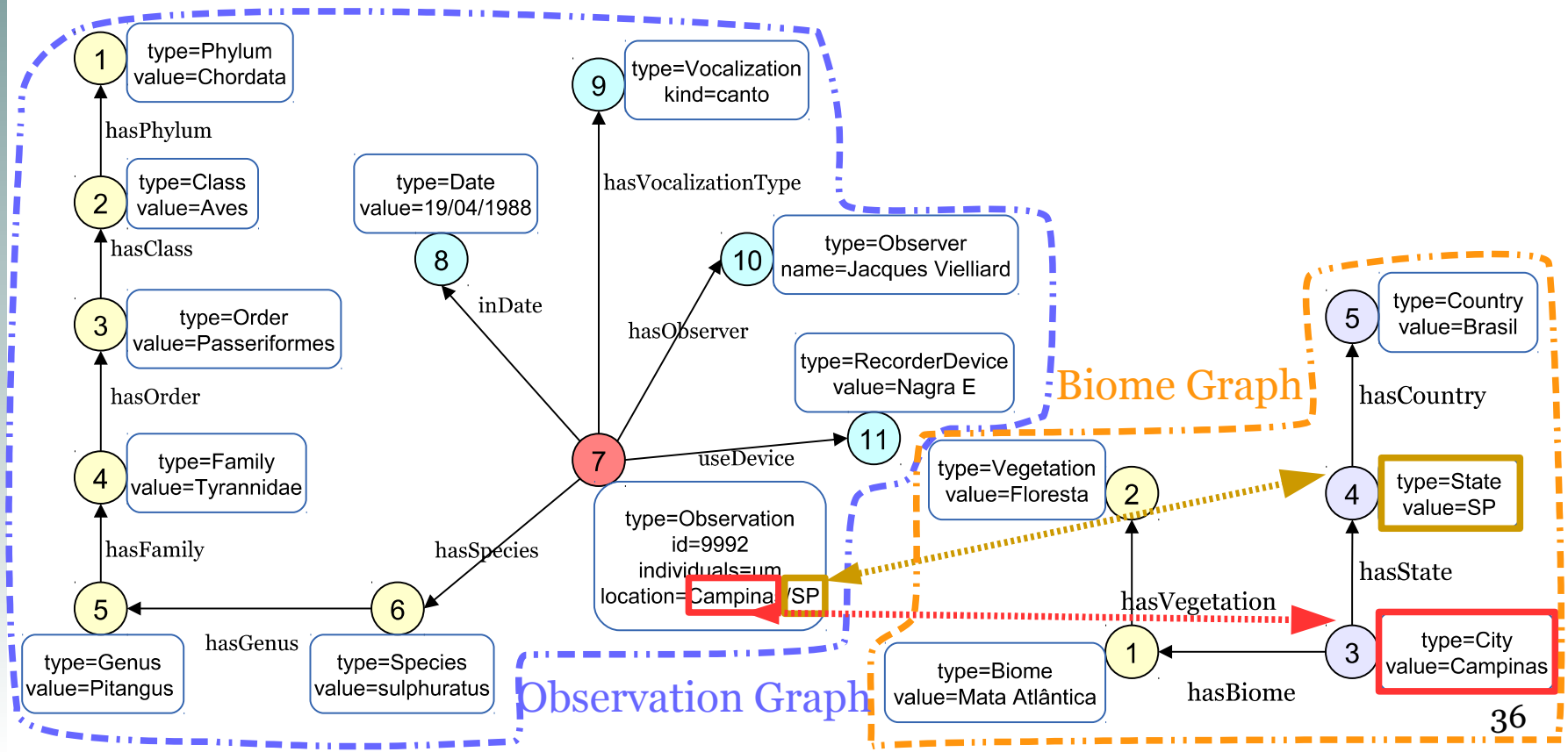
Biome Graph Database



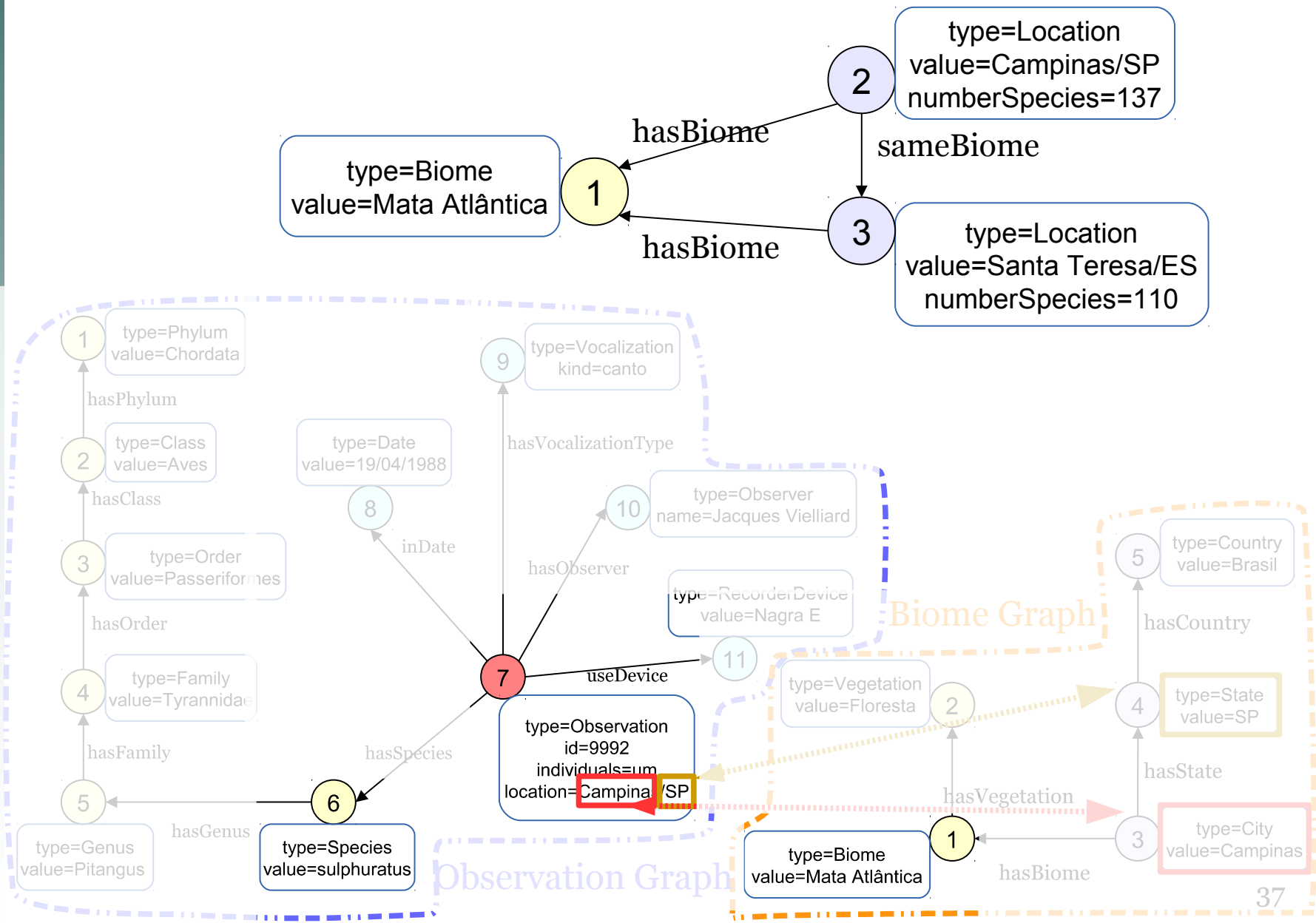
Focus 1


1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice
type=Observation
id=9992
individuals=um
location=Campinas/SP
5
type=Country
value=Brasil
hasCountry
4
type=State
value=SP
hasState
3
type=City
value=Campinas
2
type=Vegetation
value=Floresta
hasVegetation
1
type=Biome
value=Mata Atlântica
hasBiome

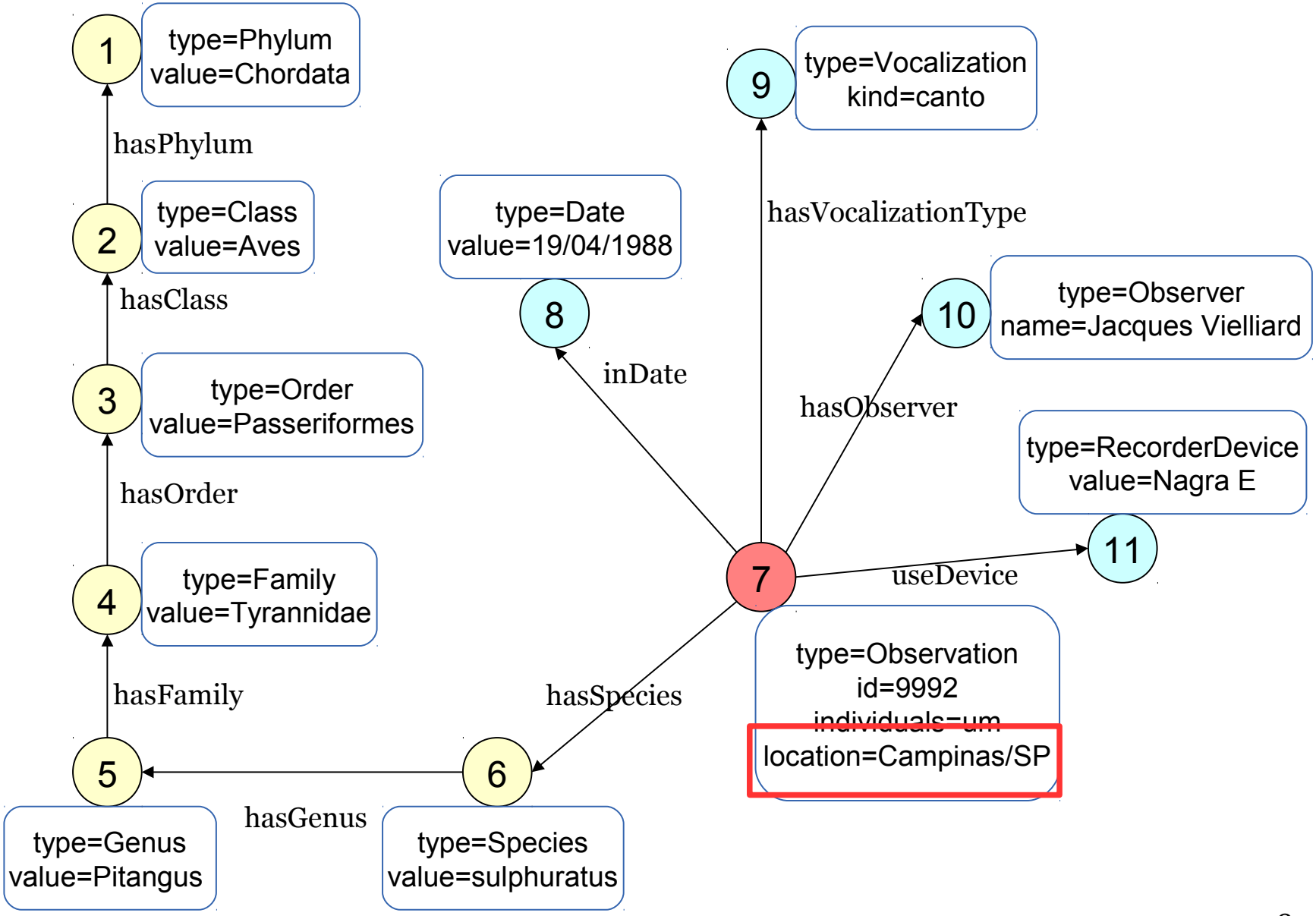
Focus 1


1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
type=Observation
id=9992
individuals=um
location=Campina /SP
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice
Observation Graph
Biome Graph
1
type=Biome
value=Mata Atlântica
2
type=Vegetation
value=Floresta
hasVegetation
3
type=City
value=Campinas
hasBiome
4
type=State
value=SP
hasState
5
type=Country
value=Brasil
hasCountry

Focus 1


type=Location
value=Campinas/SP
numberSpecies=137
2
hasBiome
sameBiome
type=Biome
value=Mata Atlântica
1
3
hasBiome
type=Location
value=Santa Teresa/ES
numberSpecies=110
1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
type=Observation
id=9992
individuals=um
location=Campinas/SP
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice
Observation Graph
type=Vegetation
value=Floresta
2
hasVegetation
type=Biome
value=Mata Atlântica
1
hasBiome
3
type=City
value=Campinas
hasState
4
type=State
value=SP
hasCountry
5
type=Country
value=Brasil
Biome Graph

Focus 1 – Step 1

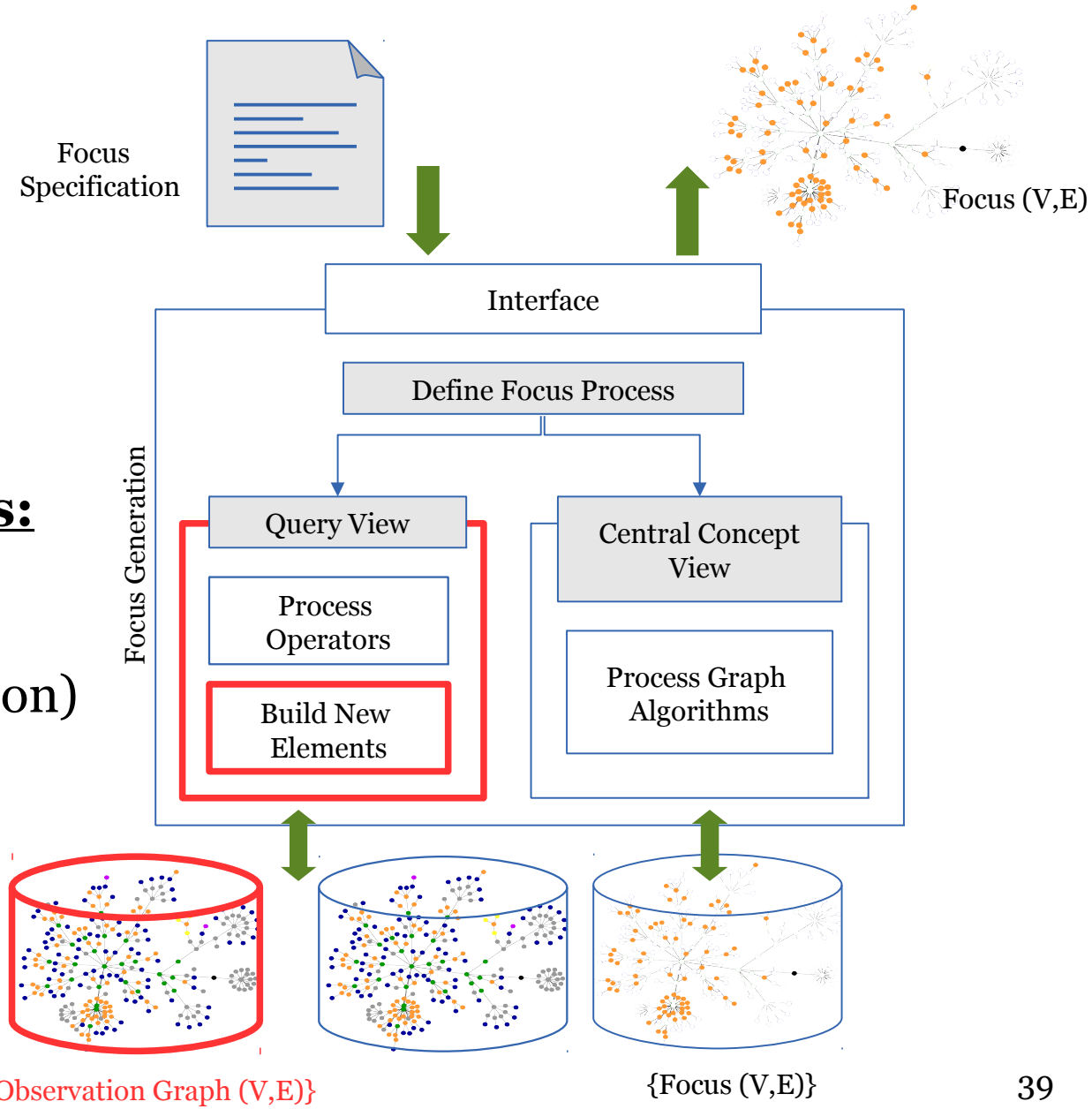


Focus 1 – Step 1

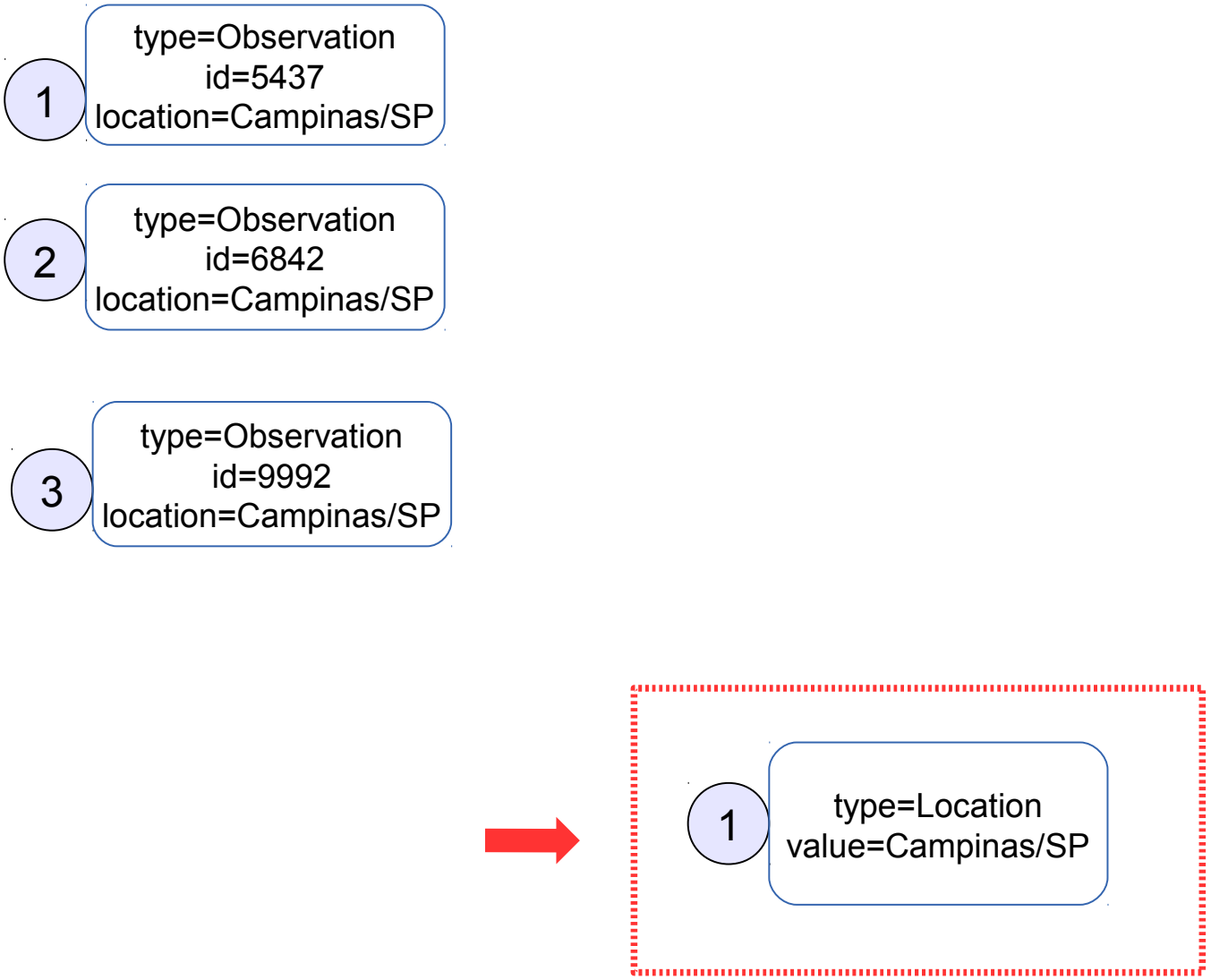
Build Elements:

V: Location

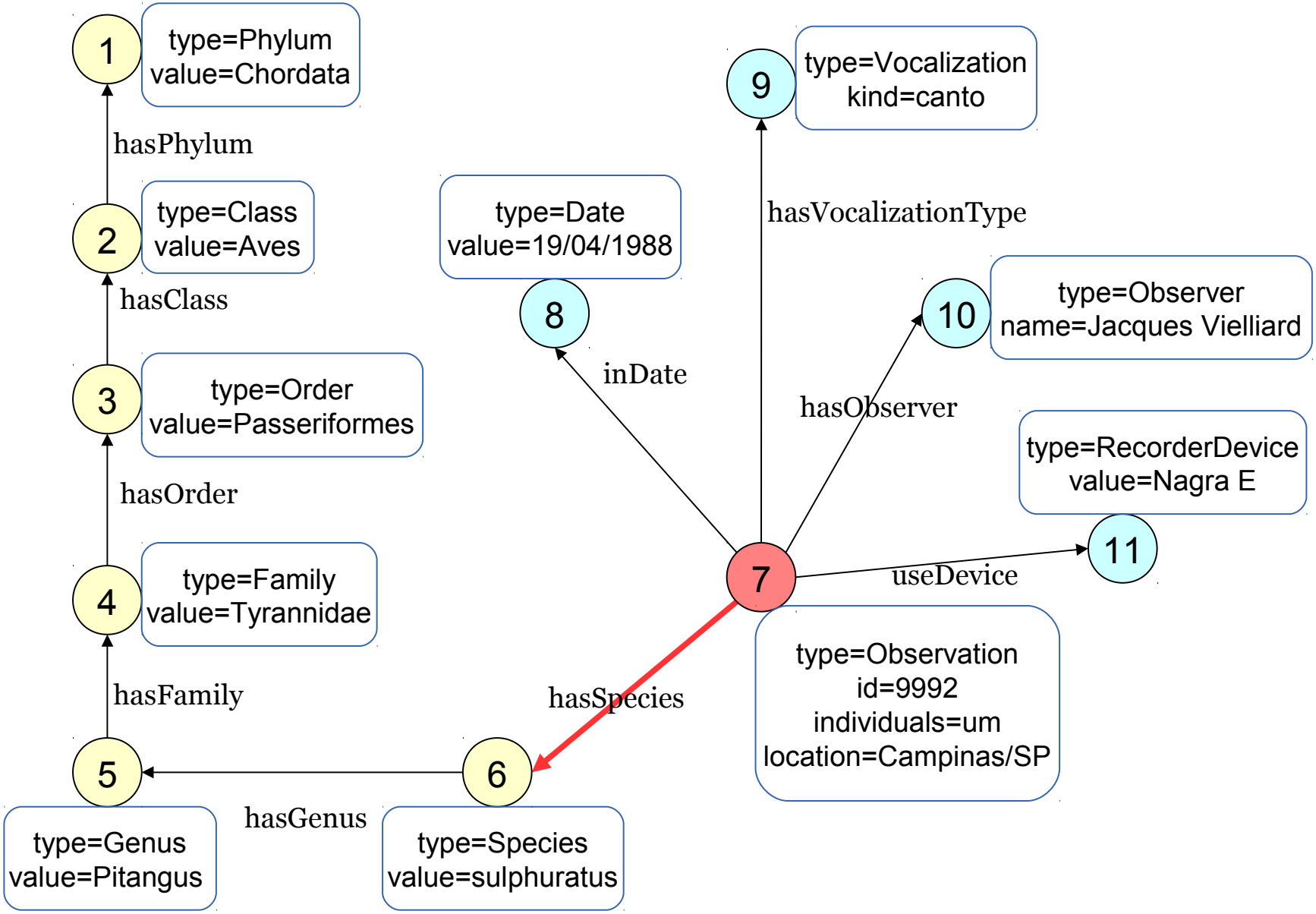
(Observation.location)



Focus 1 – Step 1



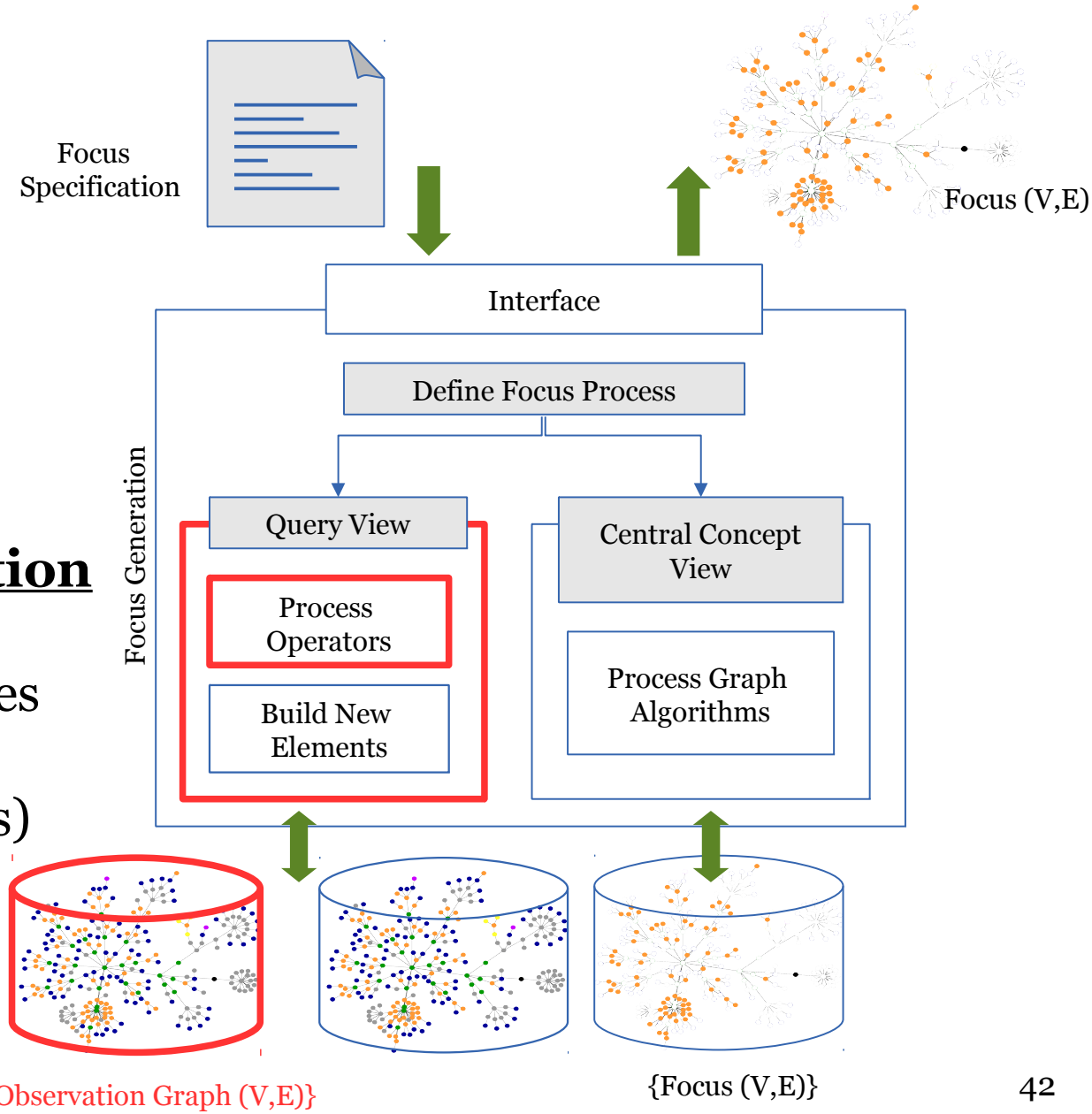
Focus 1 – Step 2



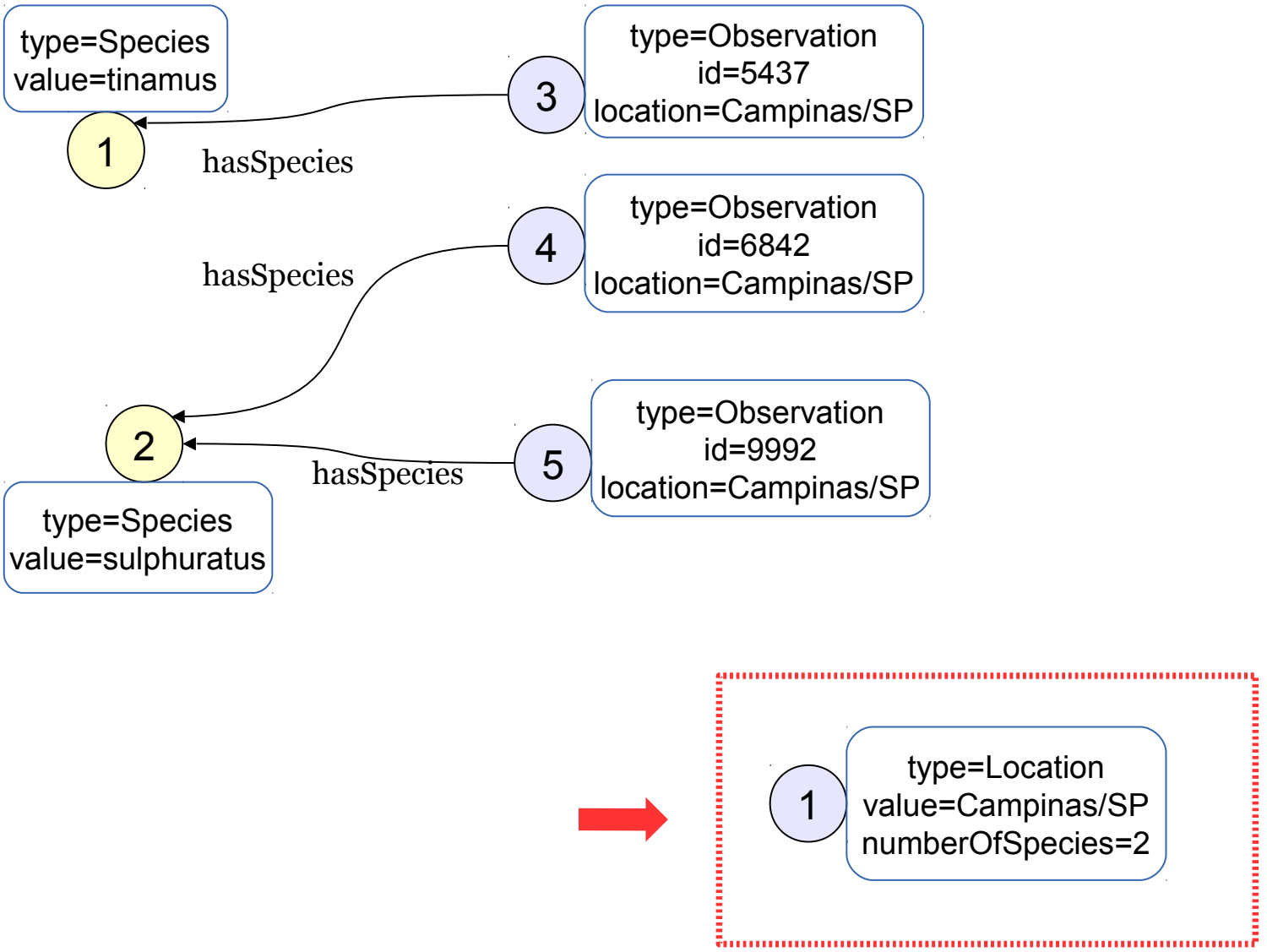
Focus 1 – Step 2

Aggregate Function

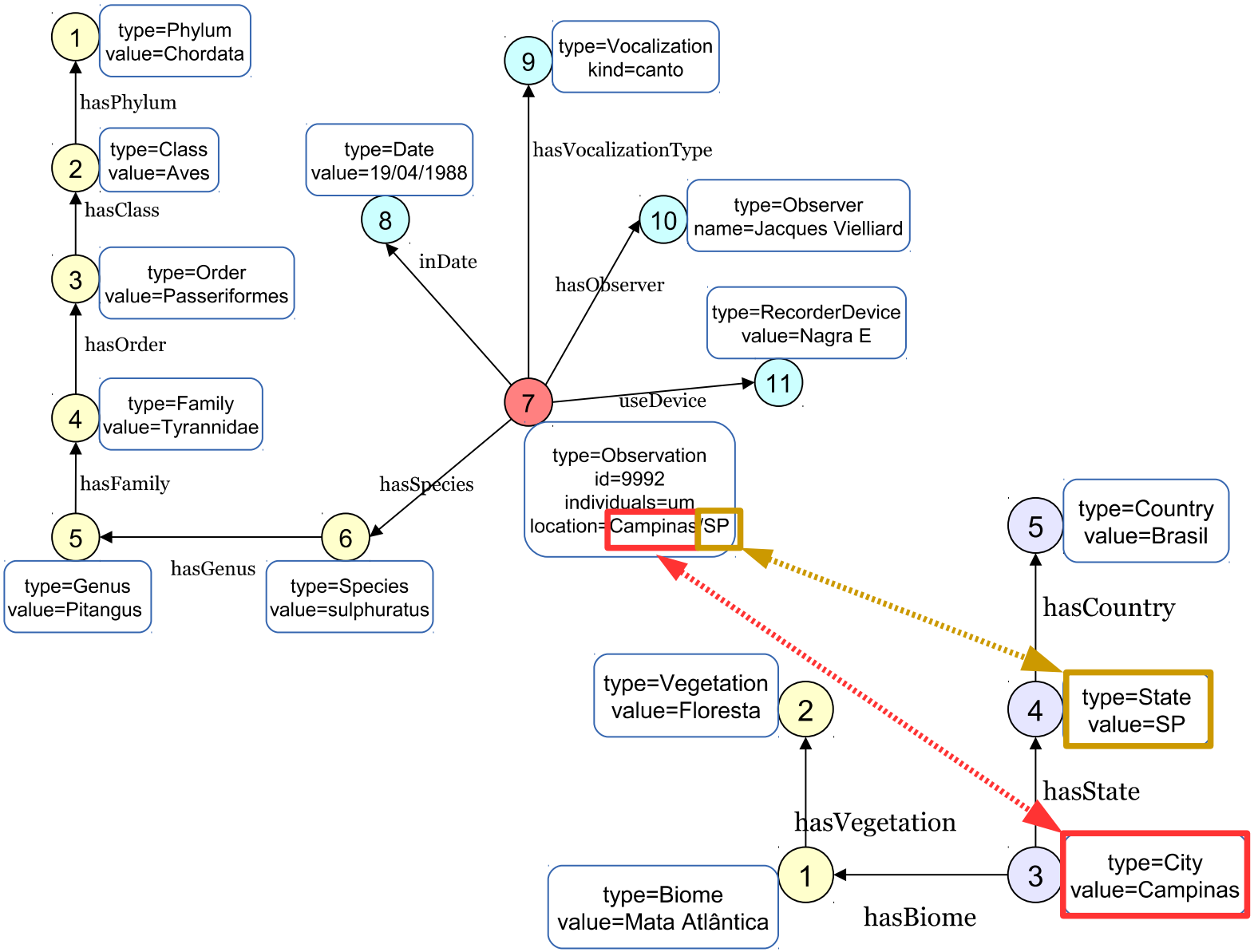
V:Location.
numberOfSpecies
= count(
distinct(hasSpecies)
)



Focus 1 – Step 2



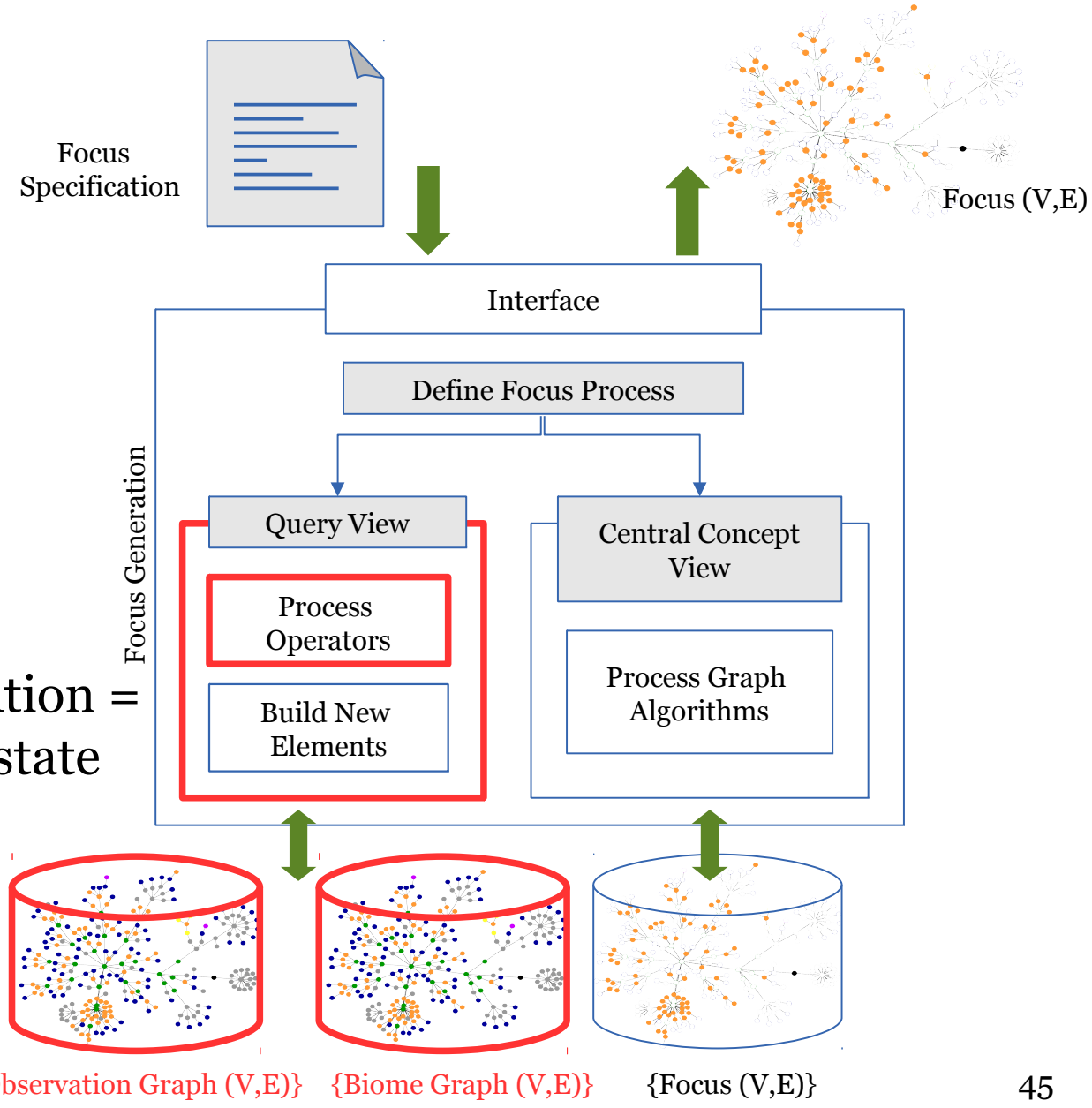
Focus 1 – Step 3



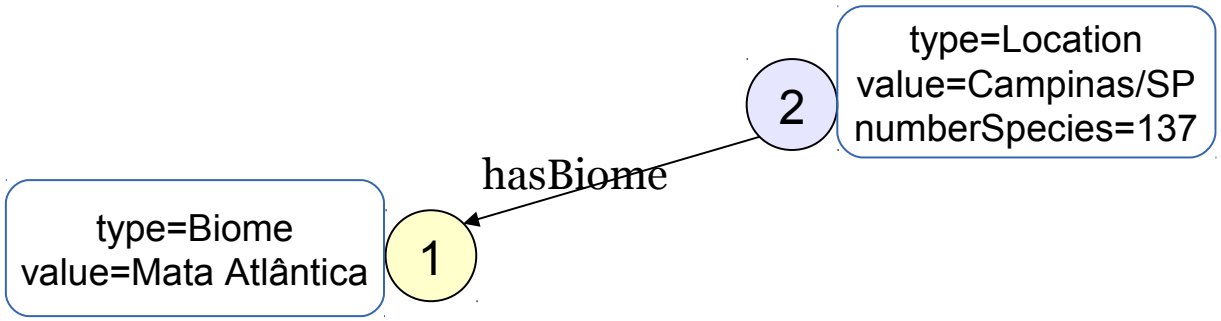
Focus 1 – Step 3

Join

E:hasBiome
V:Location, Biome
ON Observation.location =
Biome.city + Biome.state



Focus 1 – Step 3

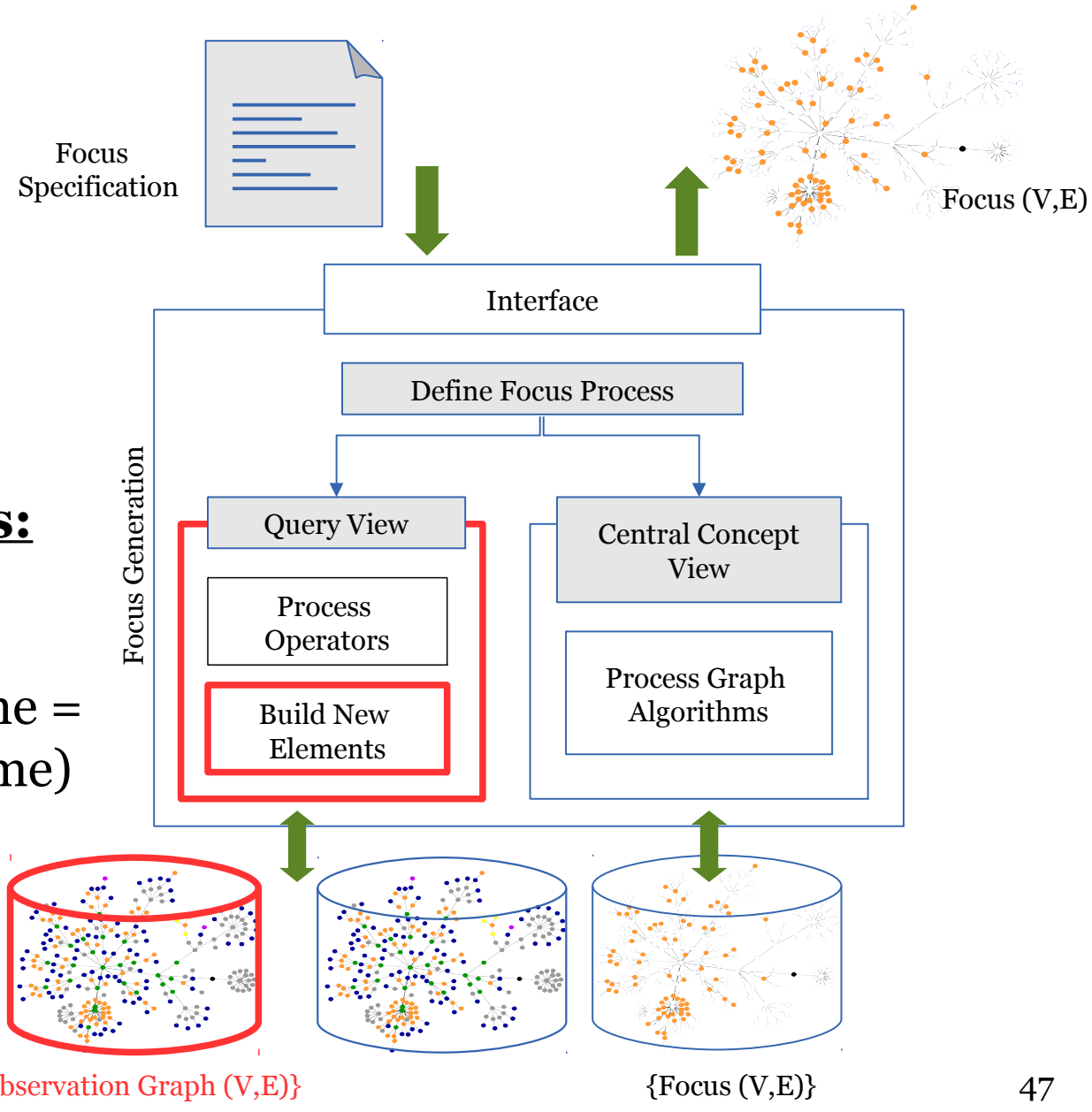


Focus 1 – Step 4

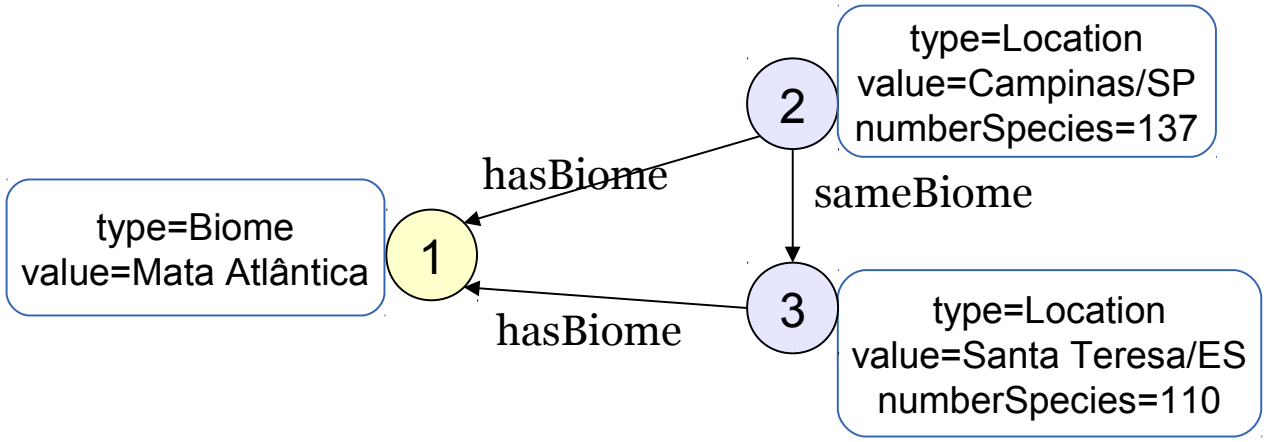
Build Elements:

E: sameBiome

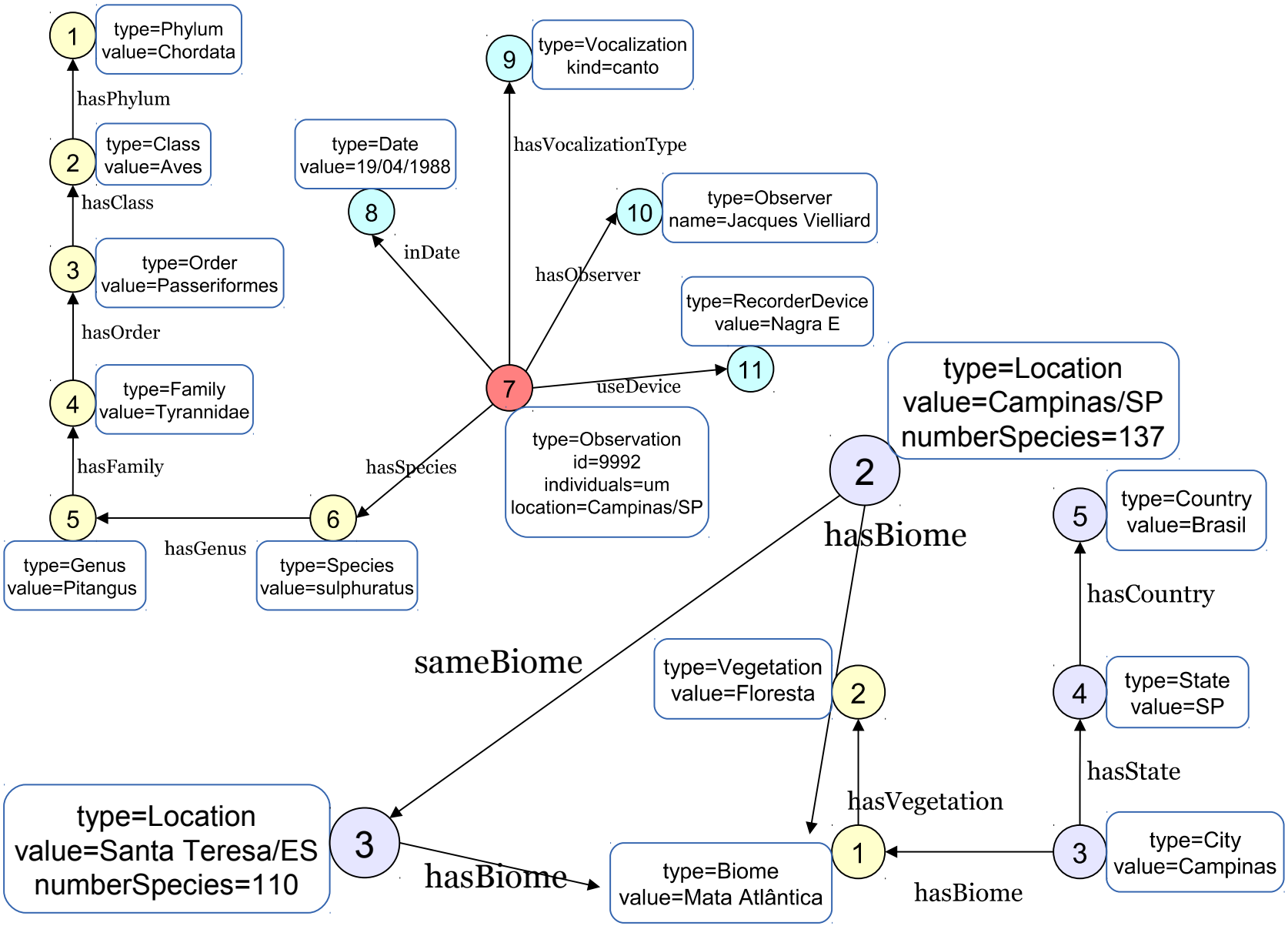
(LocationA.hasBiome = Location B.hasBiome)



Focus 1- Step 4



Focus 1- Step 4

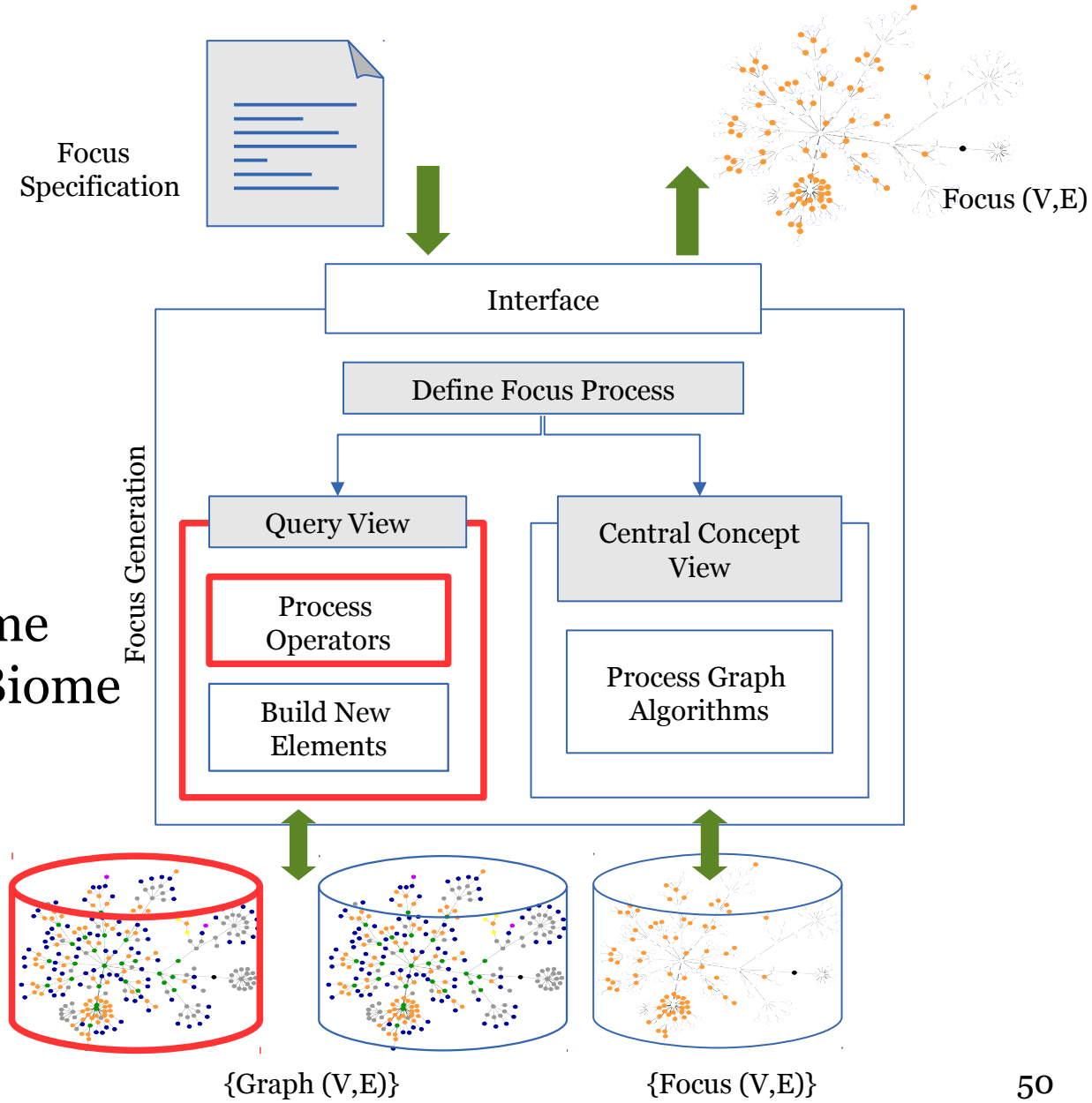


Focus 1 – Step 5

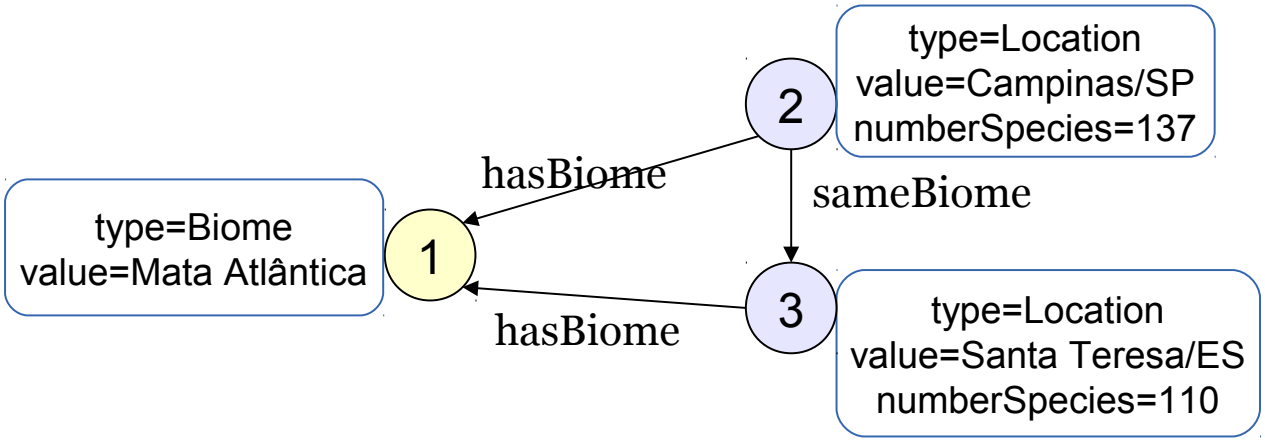
Projection

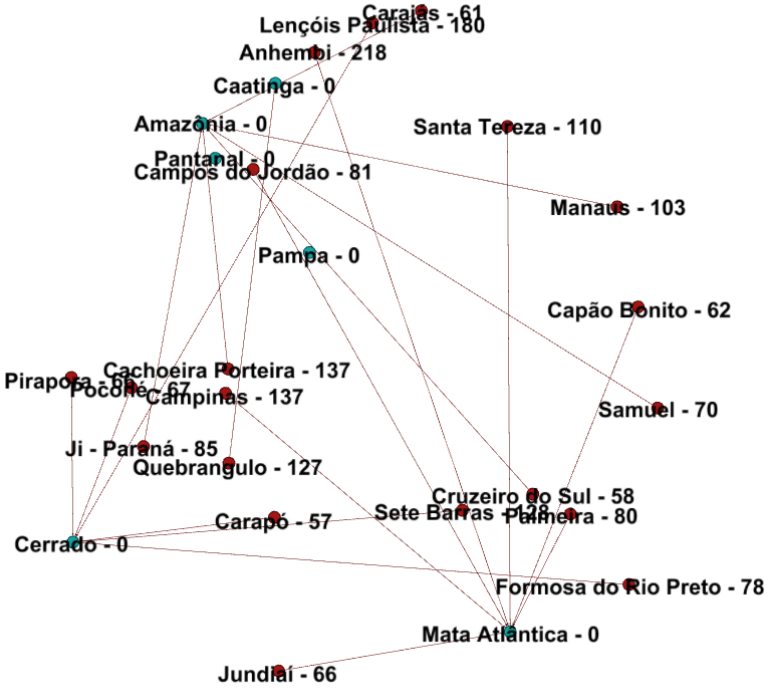
V: Location, Biome

E: hasBiome, sameBiome



Focus 1

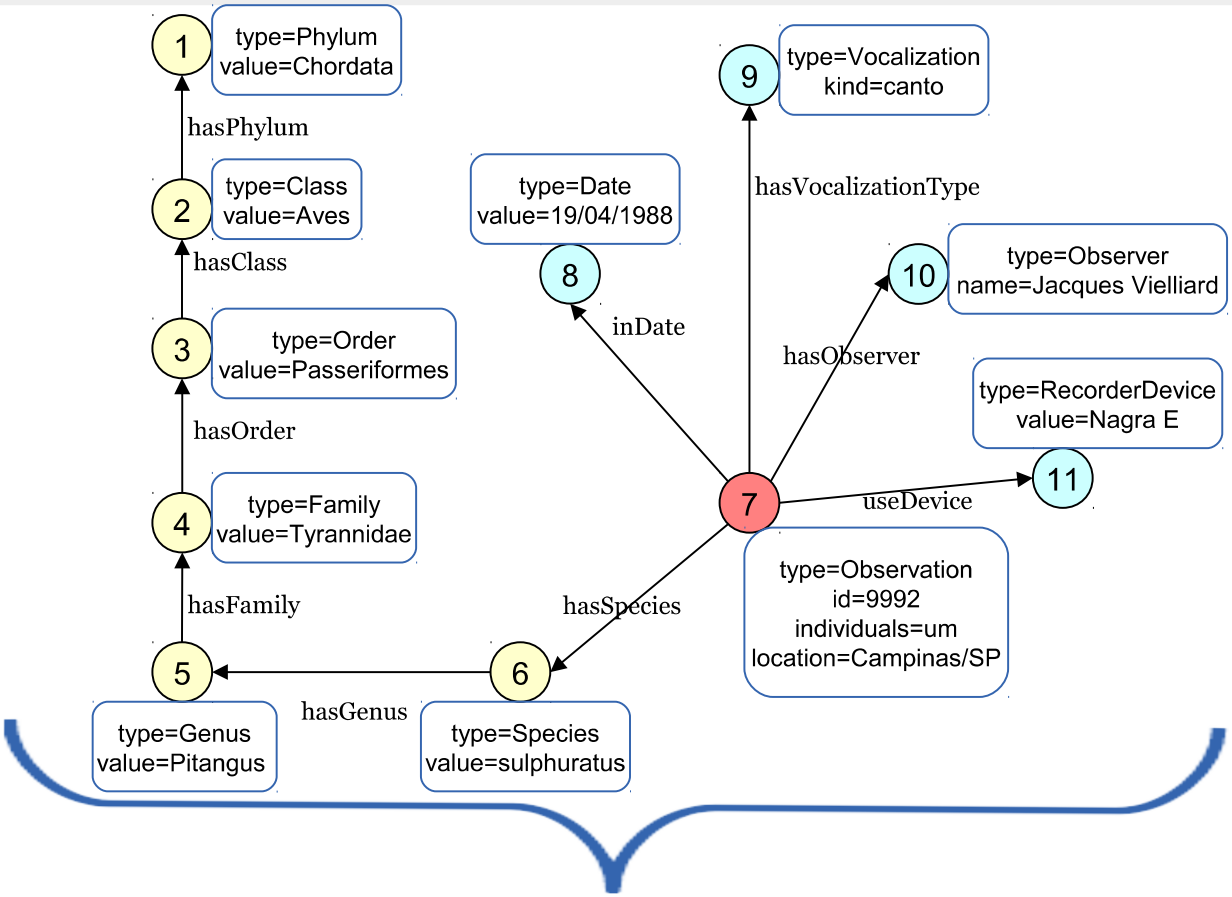

type=Location
value=Campinas/SP
numberSpecies=137
2
hasBiome
sameBiome
type=Biome
value=Mata Atlântica
1
3
hasBiome
type=Location
value=Santa Teresa/ES
numberSpecies=110


Carajás - 61
Lençóis Paulista - 180
Anhembi - 218
Caatinga - 0
Amazônia - 0
Santa Teresa - 110
Pantanal - 0
Campos do Jordão - 81
Manaus - 103
Pampa - 0
Capão Bonito - 62
Cachoeira Porteira - 137
Pirapora - 68
Pocone - 67
Campinas - 137
Samuel - 70
Ji - Paraná - 85
Quebrangulo - 127
Cruzeiro do Sul - 58
Sete Barras - 128
Palmeira - 80
Carapó - 57
Cerrado - 0
Formosa do Rio Preto - 78
Mata Atlântica - 0
Jundiaí - 66

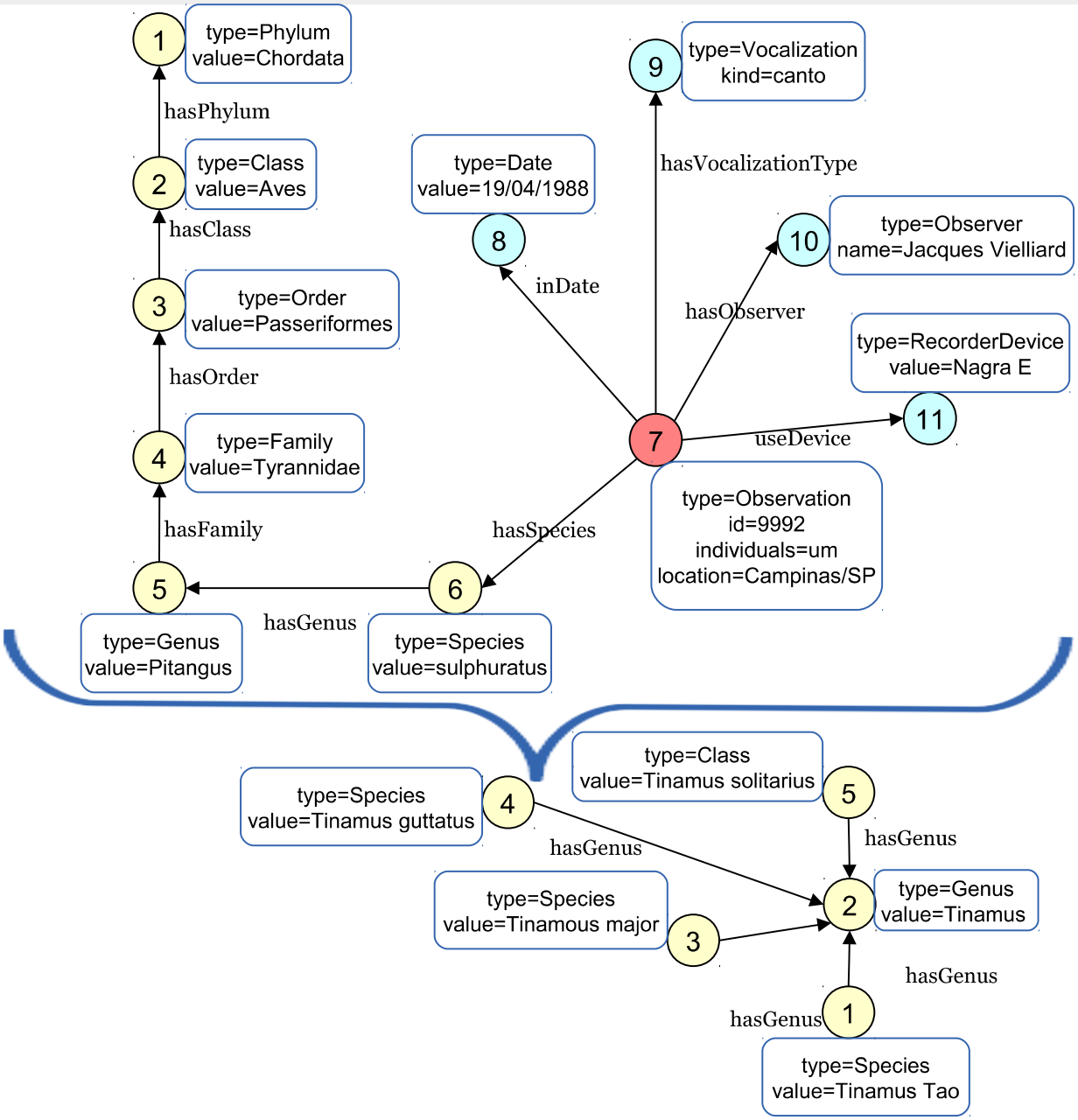
Focus 2

- *Set of species closest in the taxonomy to the species <u>Tinamus tao</u>*
  - Species
  - Genus
  - Family
  - Order
  - Class
  - Phylum

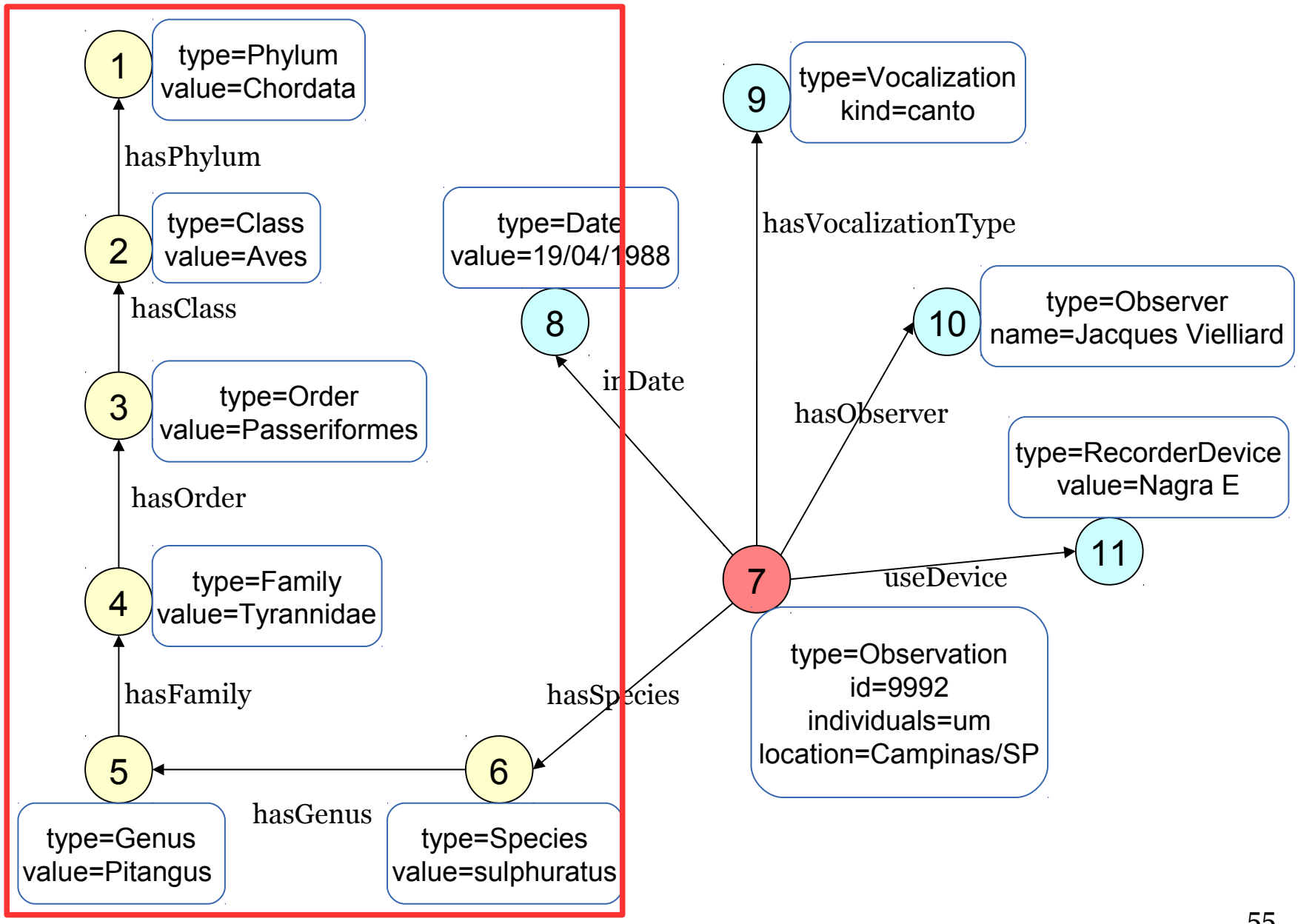
Focus 2


1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
type=Observation
id=9992
individuals=um
location=Campinas/SP
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice

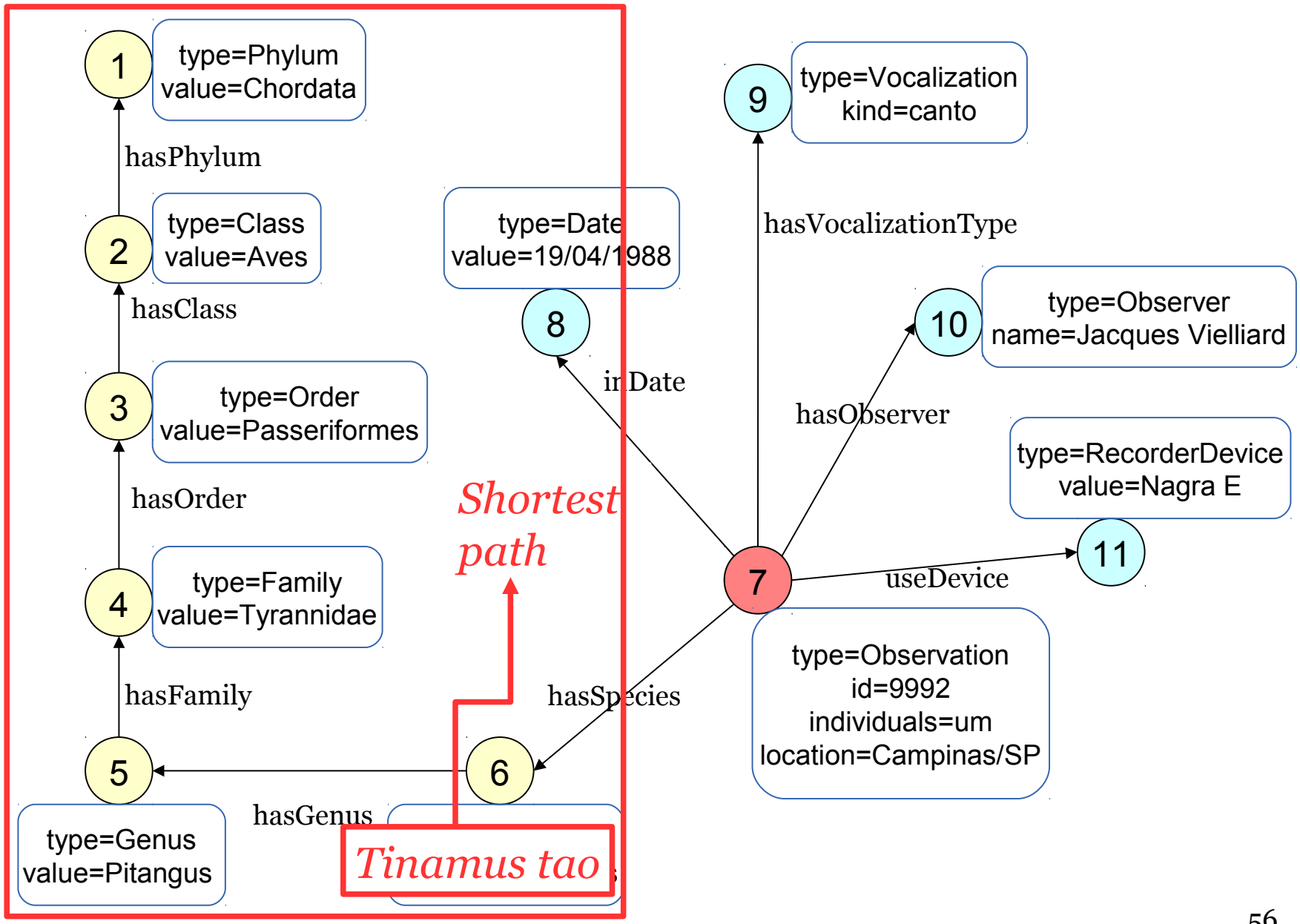
Focus 2


1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
type=Observation
id=9992
individuals=um
location=Campinas/SP
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice
4
type=Species
value=Tinamus guttatus
hasGenus
5
type=Class
value=Tinamus solitarius
hasGenus
2
type=Genus
value=Tinamus
3
type=Species
value=Tinamous major
hasGenus
1
type=Species
value=Tinamus Tao
hasGenus

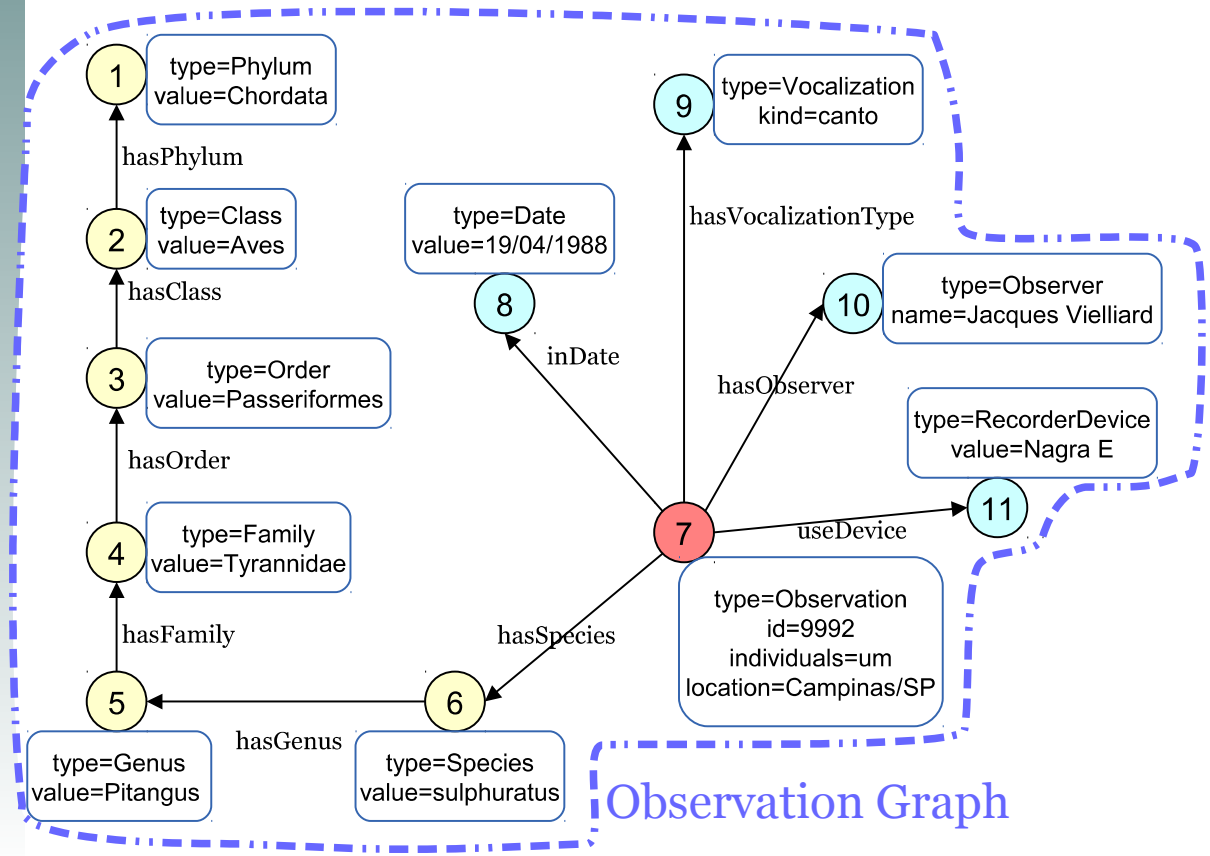
Observations Graph Database



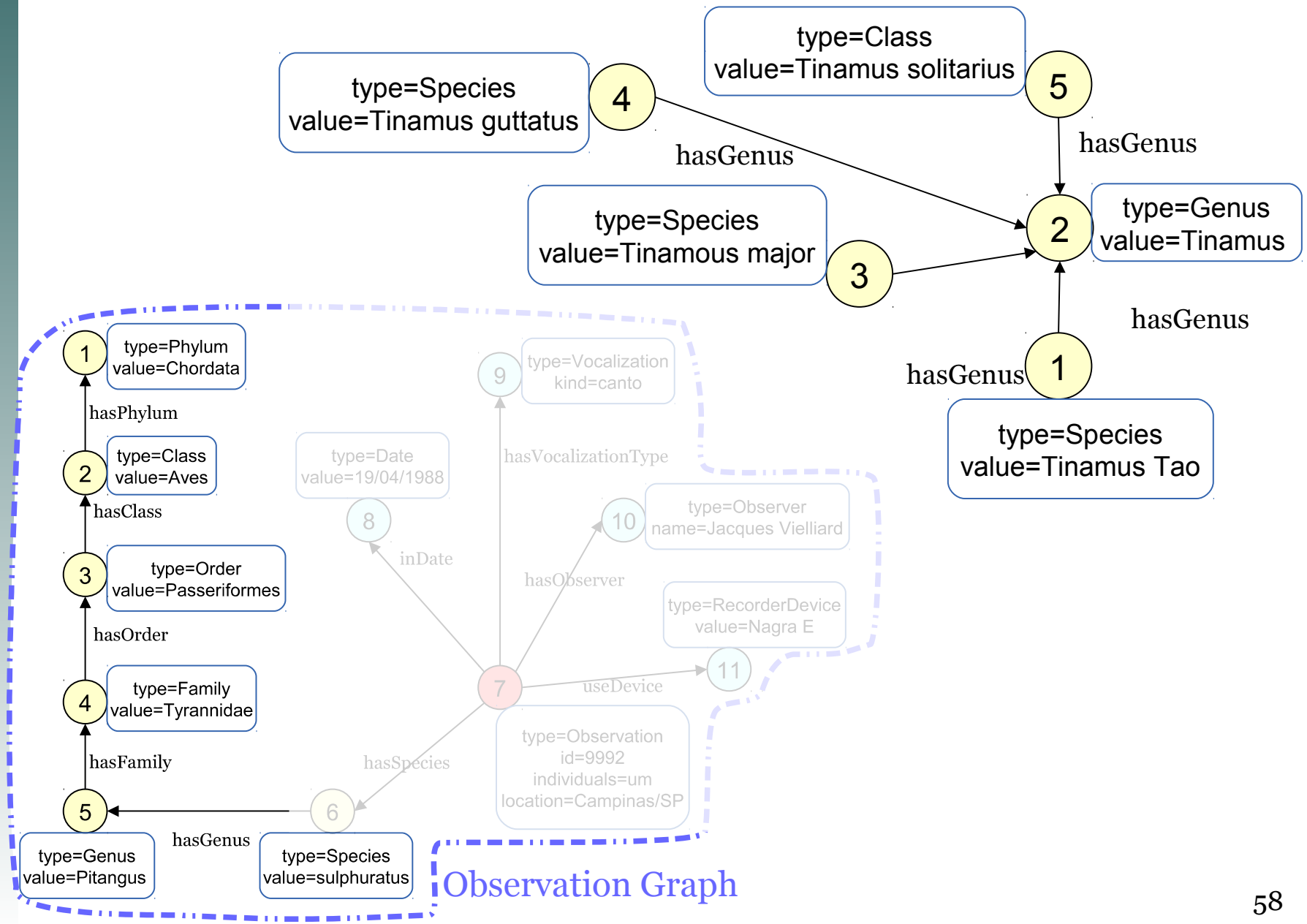
Observations Graph Database



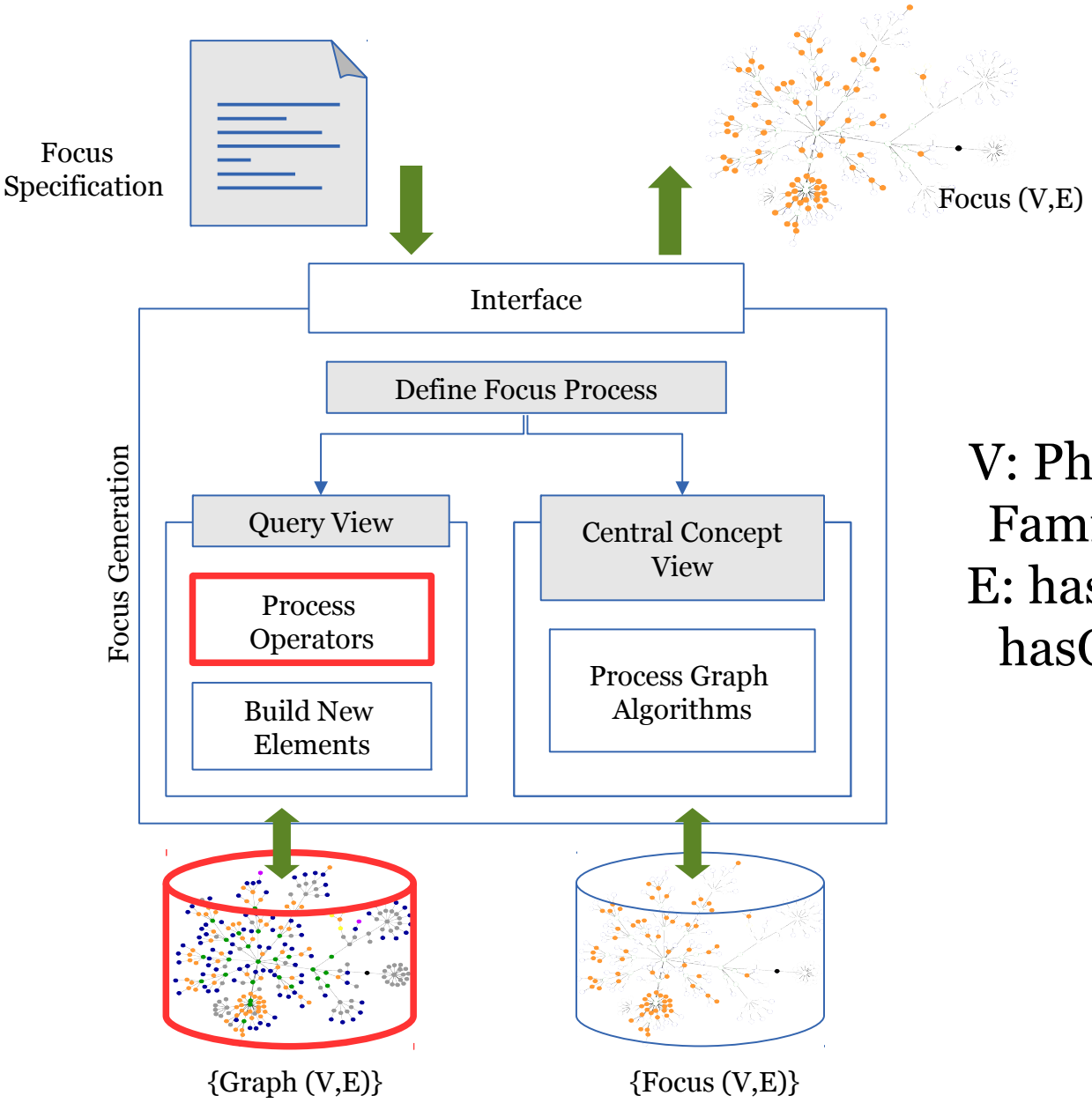
Focus 2


1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
hasSpecies
7
type=Observation
id=9992
individuals=um
location=Campinas/SP
8
type=Date
value=19/04/1988
inDate
9
type=Vocalization
kind=canto
hasVocalizationType
10
type=Observer
name=Jacques Vielliard
hasObserver
11
type=RecorderDevice
value=Nagra E
useDevice
Observation Graph

Focus 2


type=Class
value=Tinamus solitarius
5
type=Species
value=Tinamus guttatus
4
hasGenus
hasGenus
type=Species
value=Tinamous major
3
2
type=Genus
value=Tinamus
hasGenus
hasGenus
1
type=Species
value=Tinamus Tao
1
type=Phylum
value=Chordata
hasPhylum
2
type=Class
value=Aves
hasClass
3
type=Order
value=Passeriformes
hasOrder
4
type=Family
value=Tyrannidae
hasFamily
5
type=Genus
value=Pitangus
hasGenus
6
type=Species
value=sulphuratus
9
type=Vocalization
kind=canto
hasVocalizationType
type=Date
value=19/04/1988
8
inDate
10
type=Observer
name=Jacques Vielliard
hasObserver
type=RecorderDevice
value=Nagra E
11
useDevice
7
hasSpecies
type=Observation
id=9992
individuals=um
location=Campinas/SP
Observation Graph

Focus 2 – Step 1

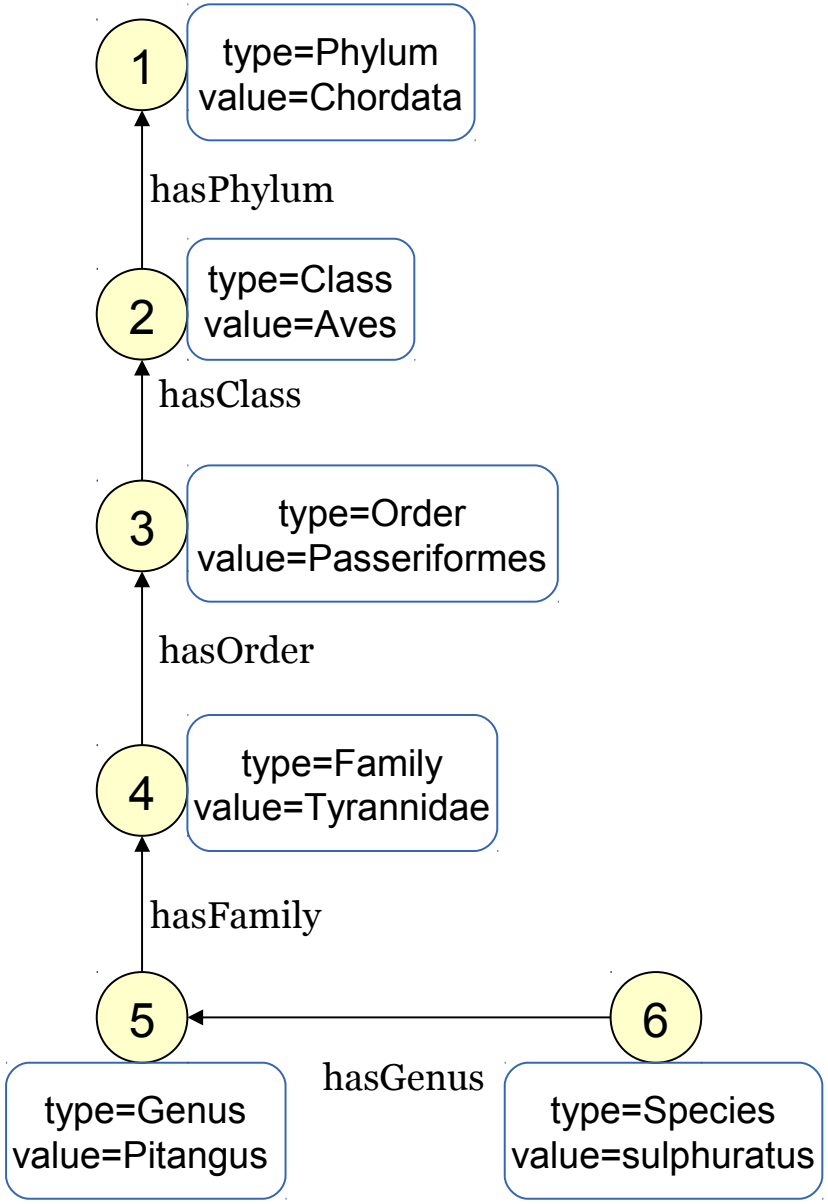


Projection

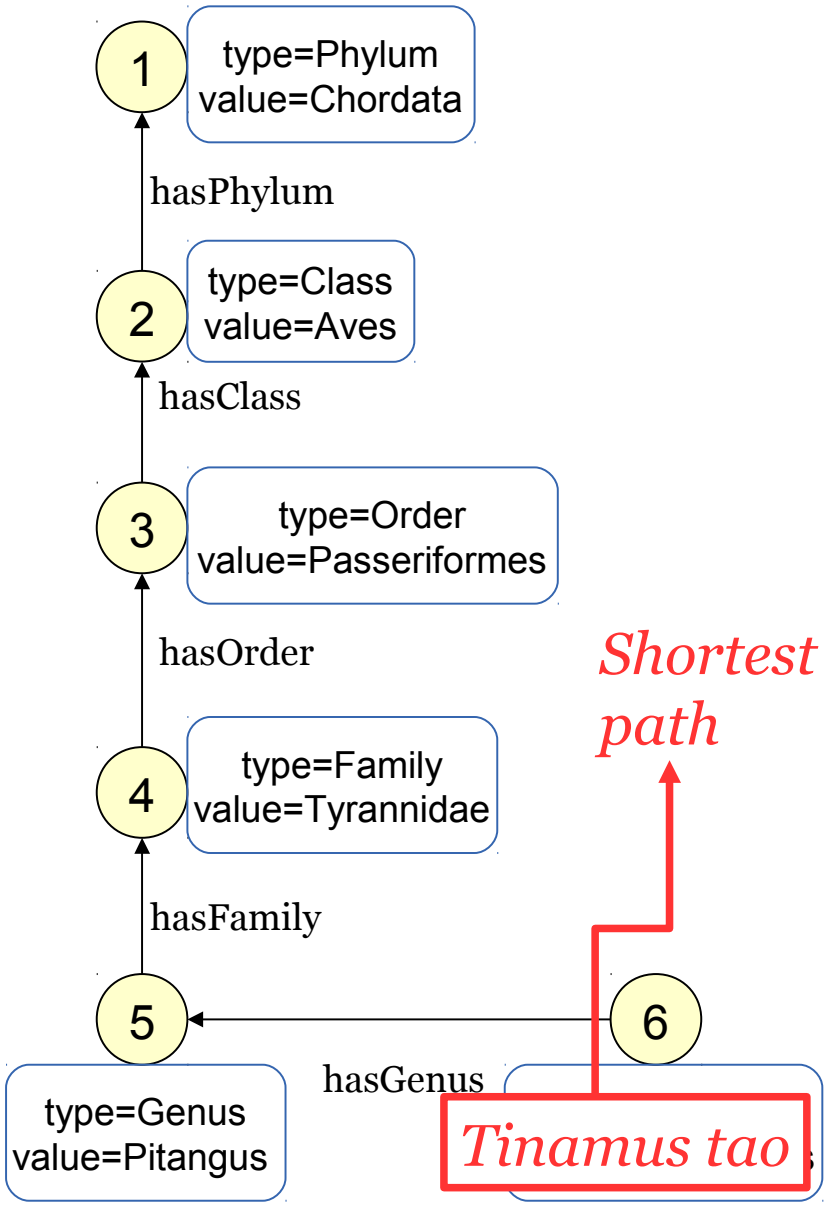
V: Phylum, Class, Order, Family, Genus, Species

E: hasPhylum, hasClass, hasOrder, hasFamily, hasGenus

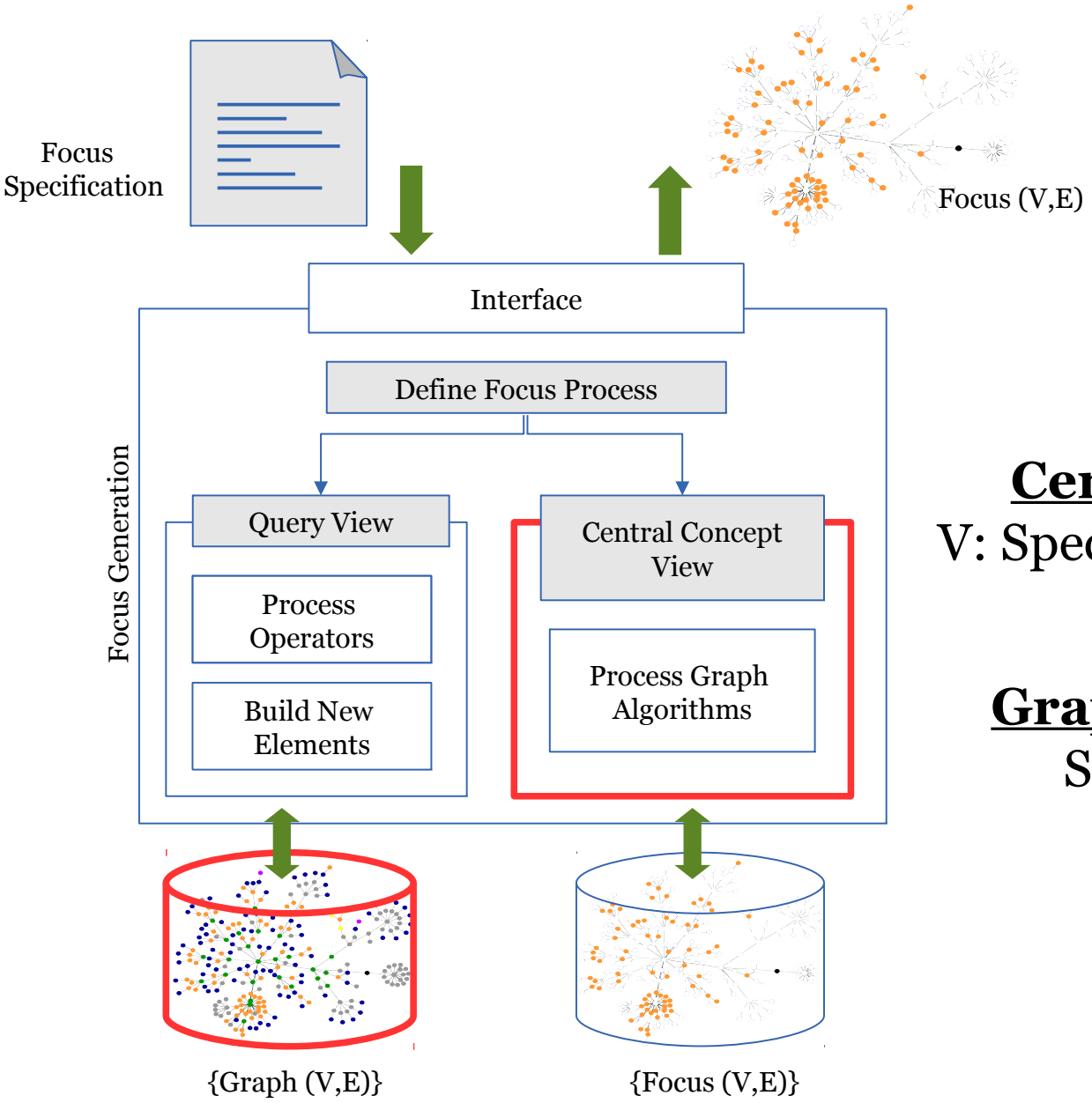
Focus 2 – Step 1



Focus 2 – Step 2



Focus 2 – Step 2



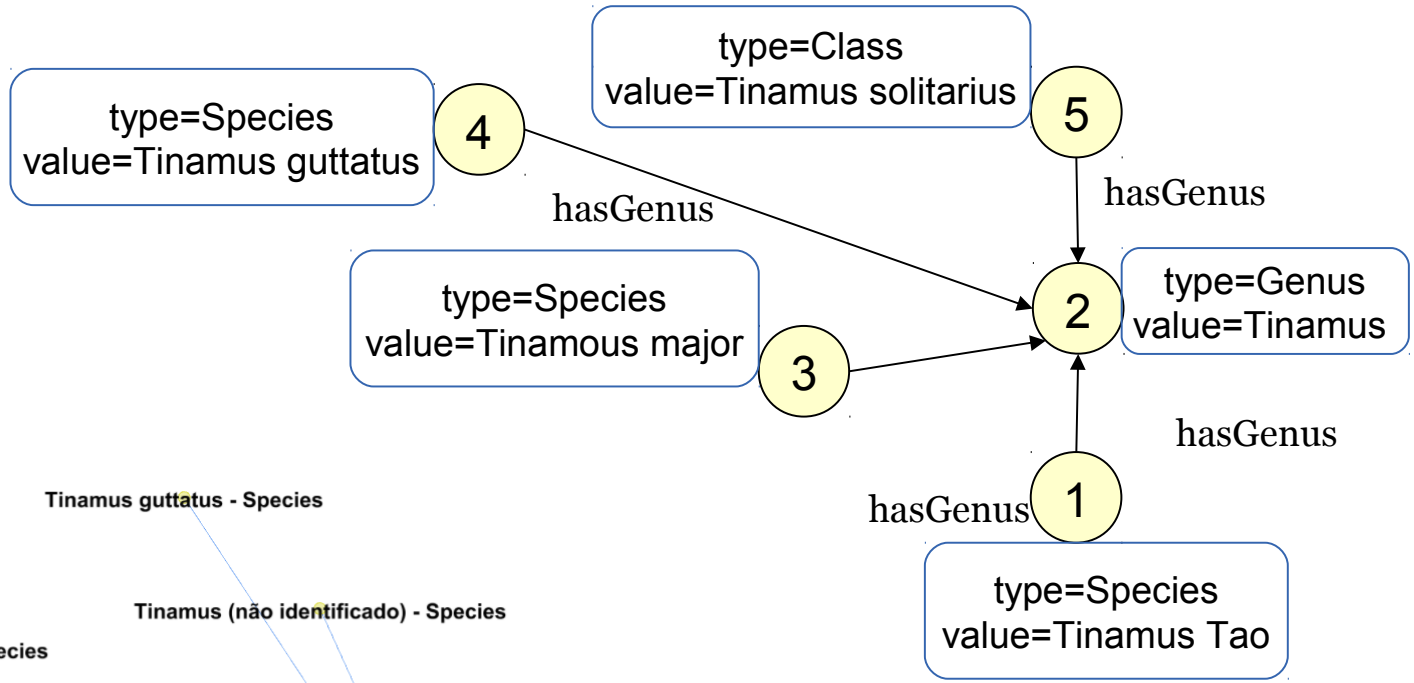
Central Concept

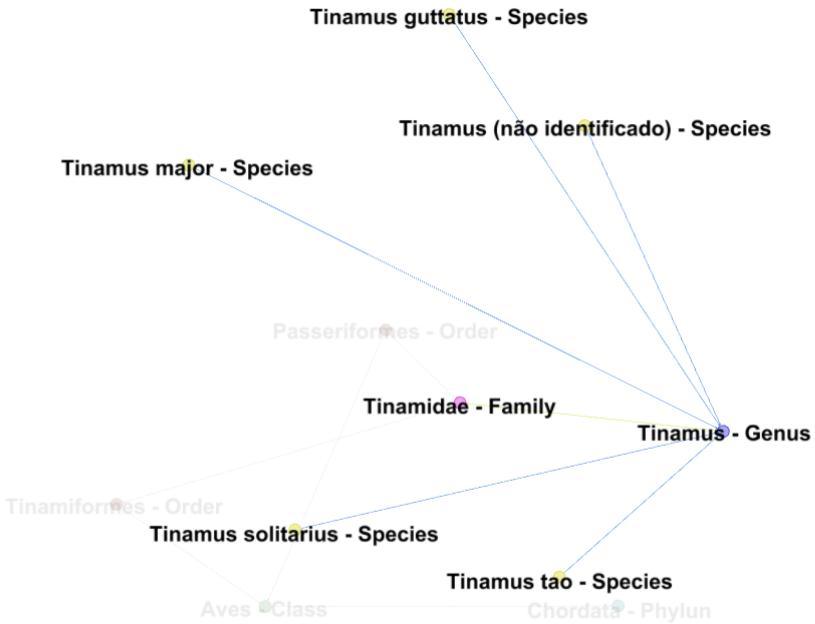
V: Species {Tinamus tao}

Graph Algorithm

Shortest Path

Focus 2


type=Species
value=Tinamus guttatus
4
type=Class
value=Tinamus solitarius
5
hasGenus
hasGenus
type=Species
value=Tinamous major
3
2
type=Genus
value=Tinamus
hasGenus
hasGenus
1
type=Species
value=Tinamus Tao


Tinamus guttatus - Species
Tinamus (não identificado) - Species
Tinamus major - Species
Passeriformes - Order
Tinamidae - Family
Tinamus - Genus
Tinamiformes - Order
Tinamus solitarius - Species
Tinamus tao - Species
Aves - Class
Chordata - Phylun

Contributions

- Adaptation of view concept
  - From relational databases
  - To graph databases
- Specification of a framework to model and create multiple foci
- New approach to multifocus research
  - Graph Databases + Views

Ongoing Work

- Validate data model for graph database
- Refine views operators
- Refine focus requirements
  - Scale, perspective
- Define focus specification

Acknowledgments

- Prof. Luis Felipe de Toledo (Biology Institute – UNICAMP)

- Laboratory of Information Systems (LIS)

- CAPES
- FAPESP
- CNPq (MuZOO Project)
- INCT
- Microsoft Research
- Embrapa

Thanks!

Laboratory of Information Systems

Jaudete Daltio

jaudete@ic.unicamp.com