

# EDIT's Common Data Model

And it's relation with  
TDWG standards





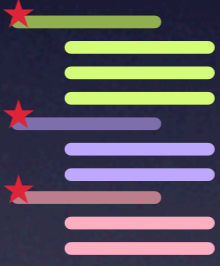
Markus Döring  
BGBM, FU-Berlin  
September 2007

# EDIT Context

- Manage data from taxonomists doing monographs, checklists, field guides or field work
- Share data between users and applications
- Integrate different applications and develop new ones
- Keep data in sync between data repositories
- *Need for a shared data model, exchange standard or at least semantic*

# TDWG Schemas

- Describe overlapping areas, no integration
- Different semantic, syntax and design patterns

DarwinCore	ABCD	TCS/SDD
		
<p>Flat, single class, no relations between attributes</p> <p>Single identifier, denormalised</p>	<p>Hierarchical, multiple classes, but sometimes hard to recognise</p> <p>Single identifier, denormalised</p>	<p>Hierarchical, multiple classes. Also embedded classes lacking ID &amp; normalisation</p> <p>Multiple identifier, normalised</p>

# Why isn't the ontology enough?

# Why isn't the ontology enough?

- TDWG ontology covers most data elements, but
  - lacks cardinalities, seeks loosely coupled data exchange

# Why isn't the ontology enough?

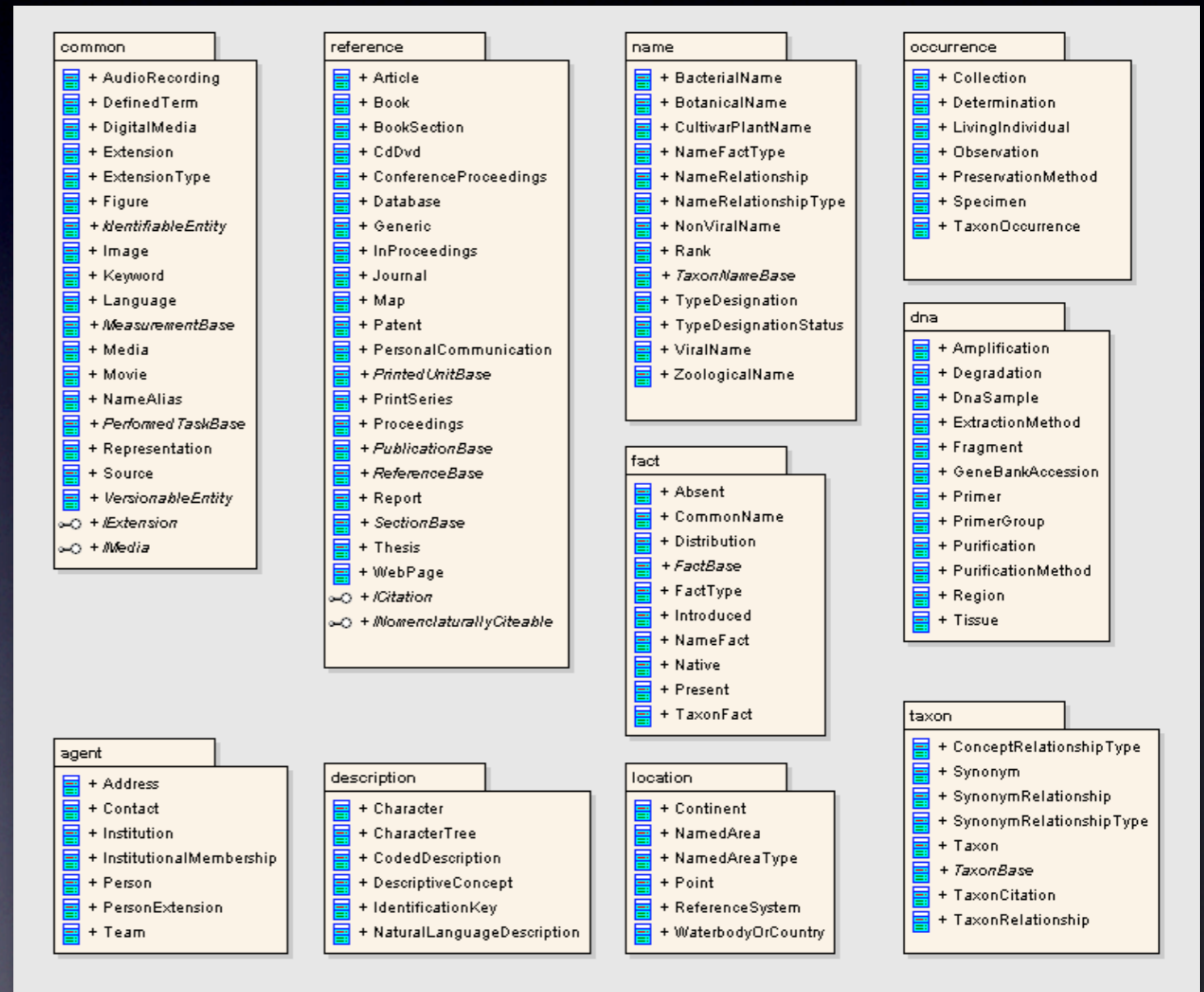
- TDWG ontology covers most data elements, but
  - lacks cardinalities, seeks loosely coupled data exchange
- Implementation model providing
  - business logic & constraints
  - strongly typed data & interfaces
  - operations
  - inheritance helpful

# Common Data Model (CDM)

- UML class model in Enterprise Architect
  - generates Java source code
- Semantically aligns primarily with TDWG ontology.
  - but other standards like BibTex, SDD, DarwinCore considered
  - naming of classes & attributes
  - source for defined terms with identifiers. E.g. name ranks
  - intention to feed back new developments into ontology
- Roadmap
  - 4Q/2007 freeze model
  - 1Q/2008 release library and basic store

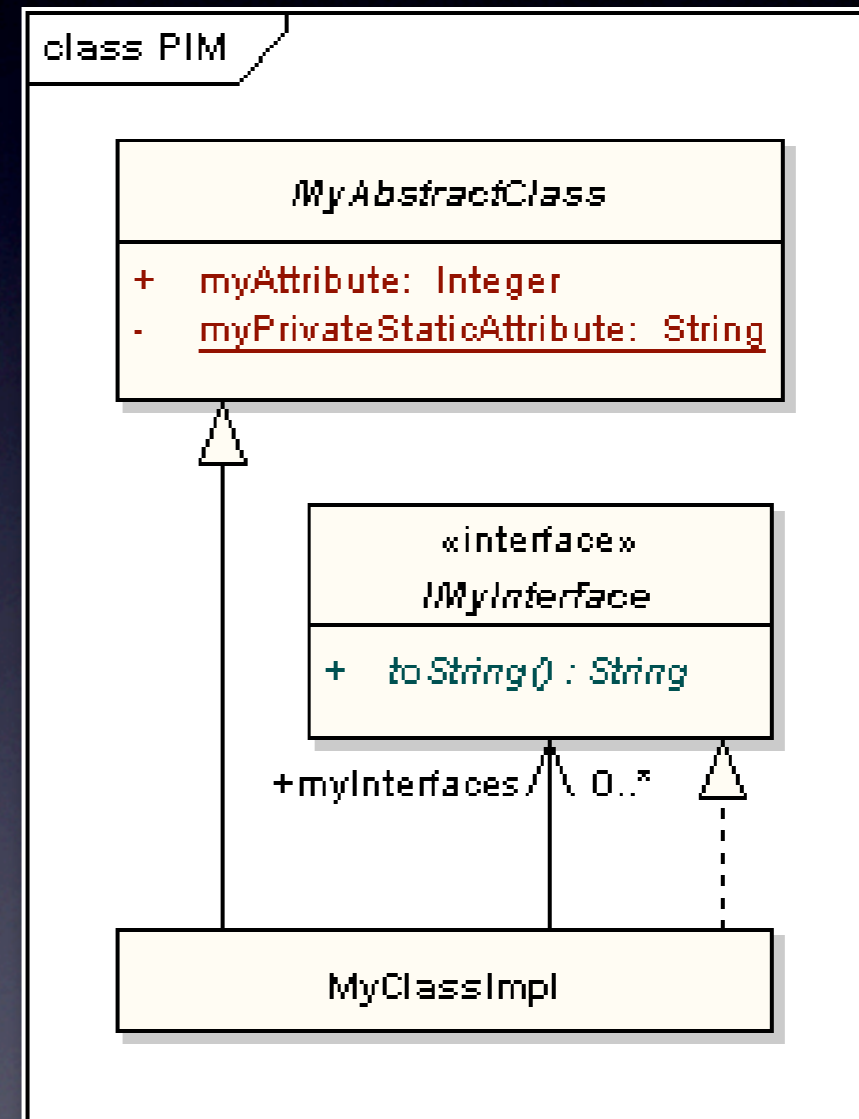
# CDM Packages / Scope

- Common
- Reference
- Agent
- Name
- Taxon
- Location
- Occurrence
- Description
- Fact
- DNA



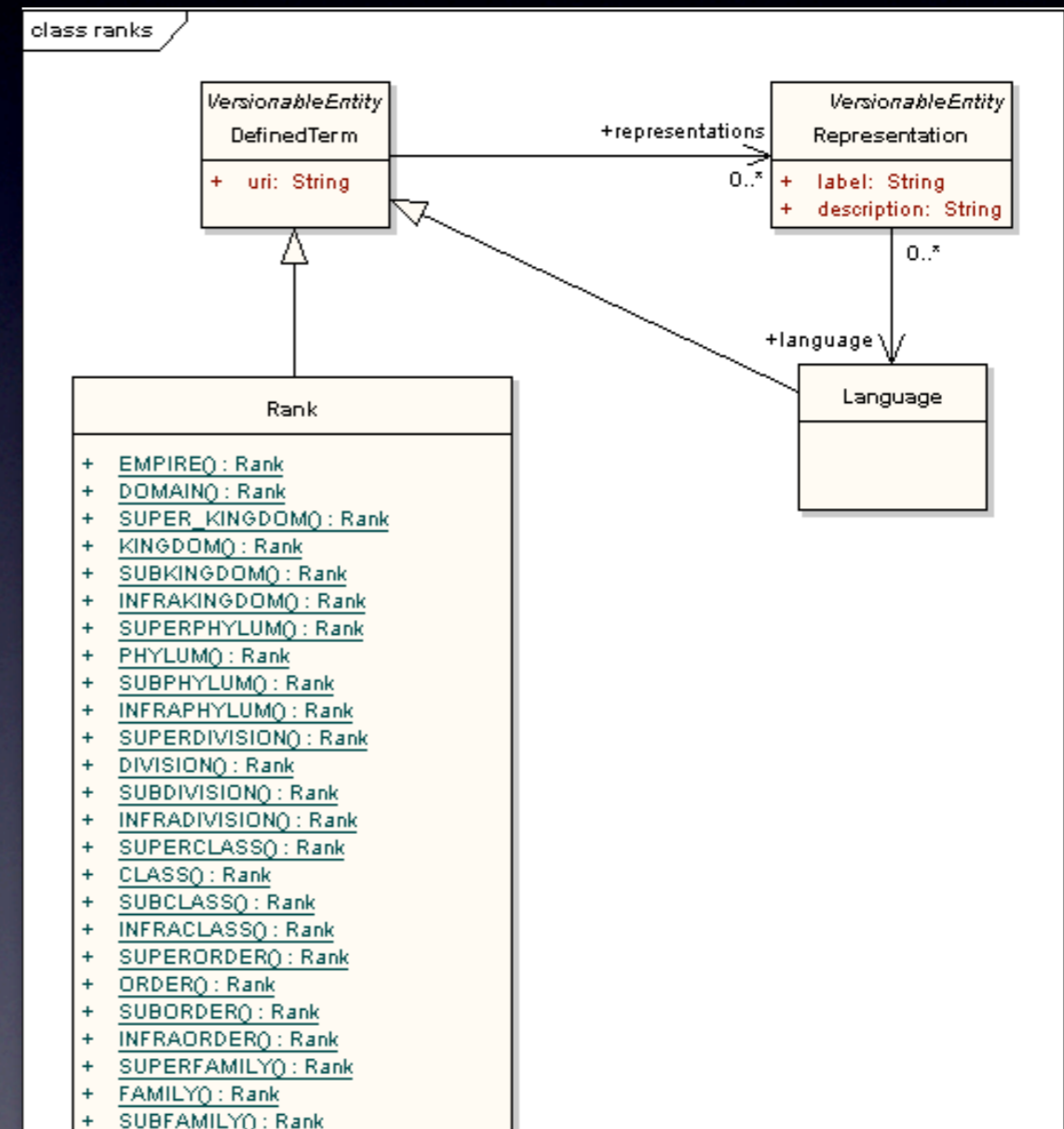
# UML Class Diagrams

- Classes
  - attribute w/ datatype
- Classes & Interfaces
  - operation w/ parameters incl properties
- Relations
  - generalisation
  - association w/ cardinality
  - realisation



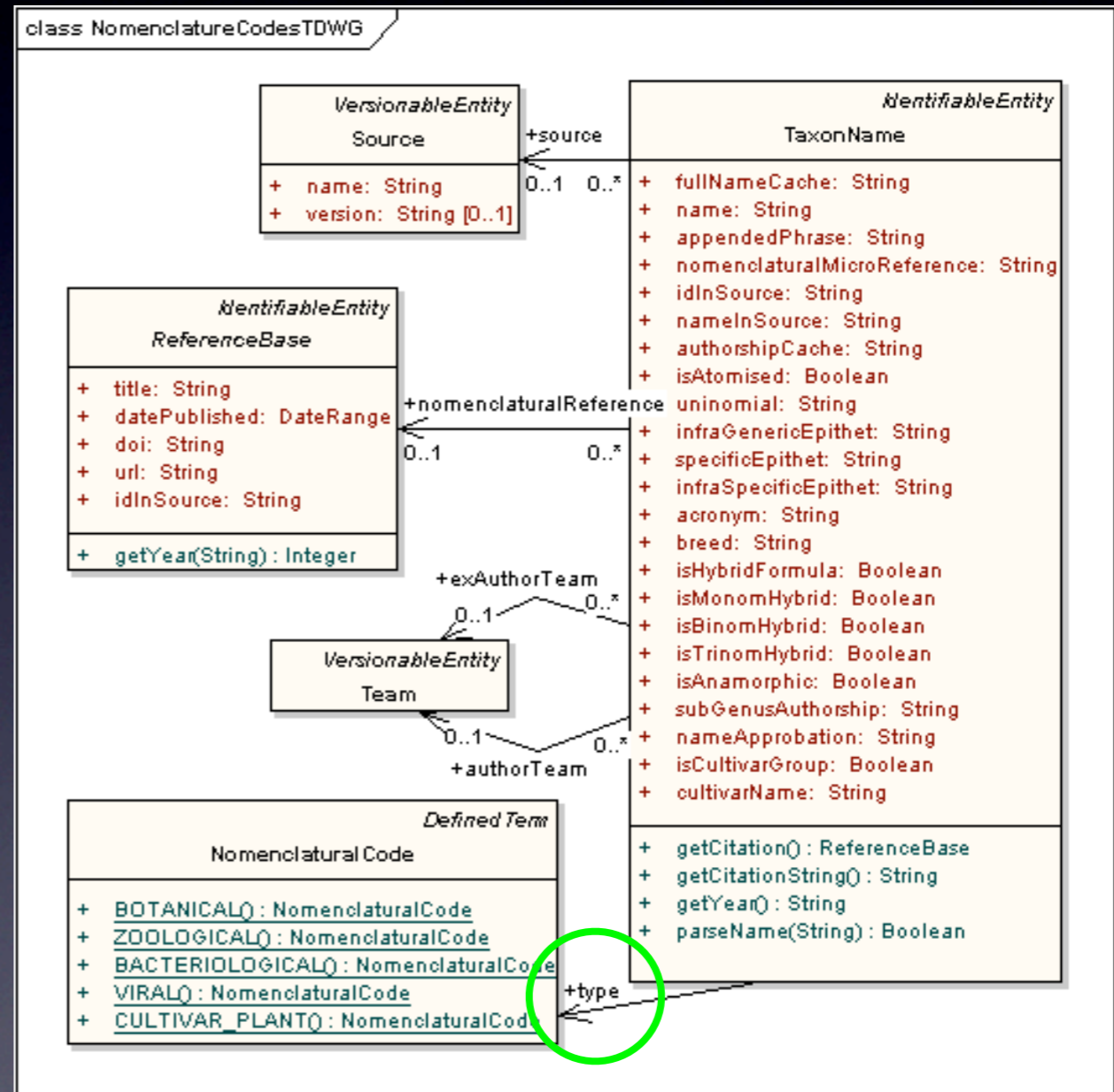
# Defined Terms

- **Extendible, persistent**
  - but common terms as static methods
- **Ontology URIs**
  - in external files, not UML
- **Multiple languages**
  - ok for defined terms
  - but is a description in another language still the same?



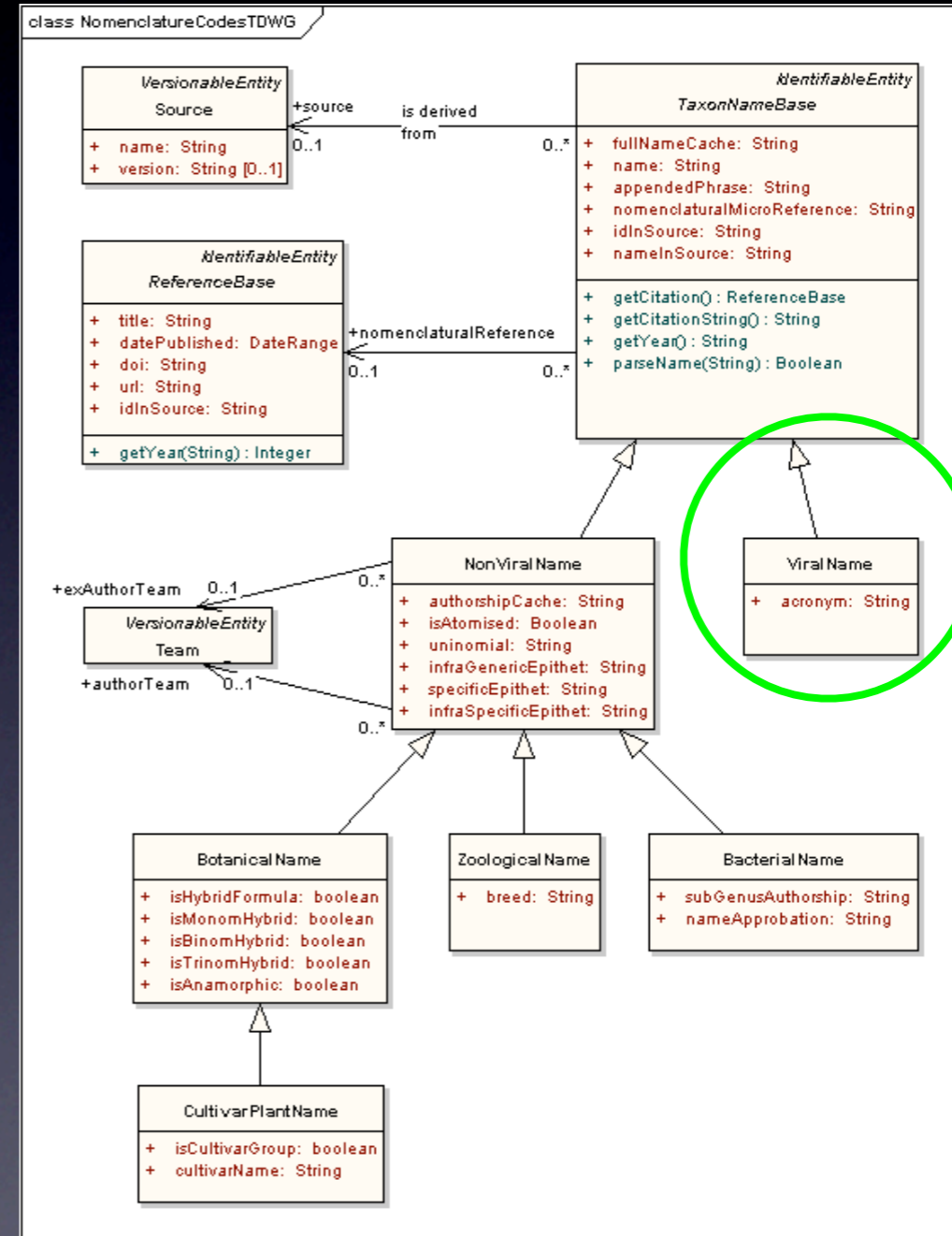
# Inheritance

- Type attribute
- extend dynamically at runtime
- CDM: Name & Taxon Relationship Type



# Inheritance

- Type attribute
  - extend dynamically at runtime
  - CDM: Name & Taxon Relationship Type
- Sub classes
  - diff. attributes+behaviour
  - fixed set of classes
  - CDM: Publication Type NomenclaturalCode



# Shared Vocabularies

- Vital for data integration
  - even if not all terms are covered
- CDM terms as *defined terms*
  - ranks, regions, countries, type designation & occurrence status, collection & institution types, name & taxon relationship types
- CDM terms as *classes*
  - nomenclatural code, reference type
- Still needed
  - basic characters or species profile model categories
  - applicability statement: ISO languages & countries, sea regions

# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



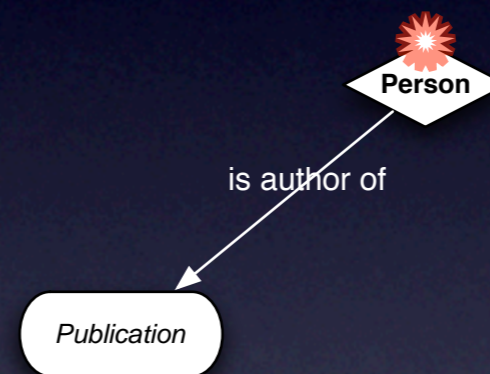
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



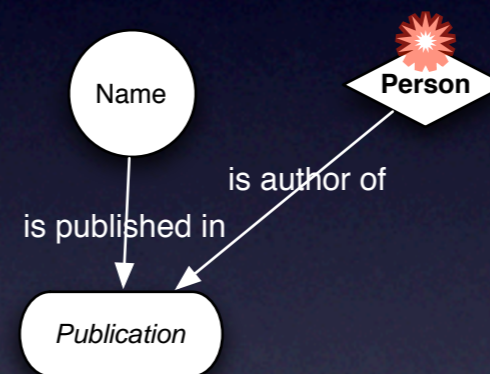
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



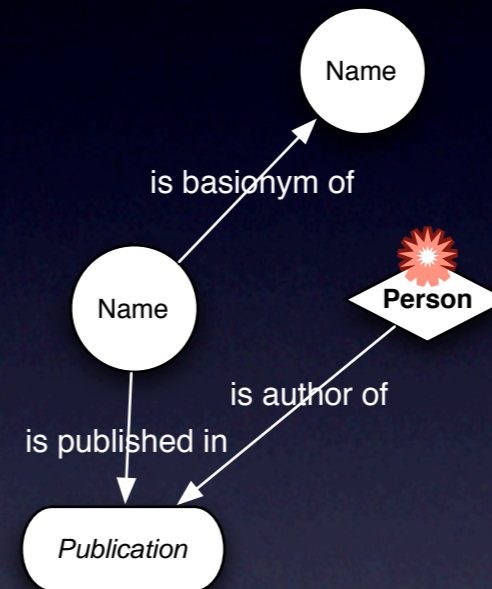
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



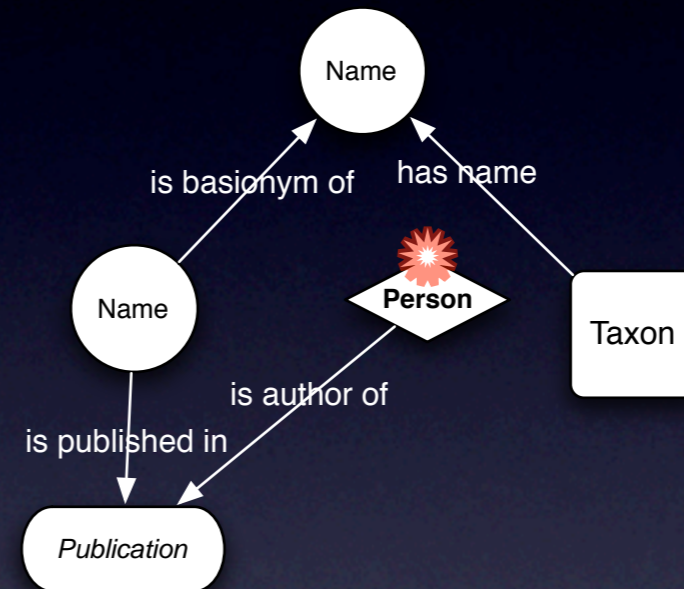
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



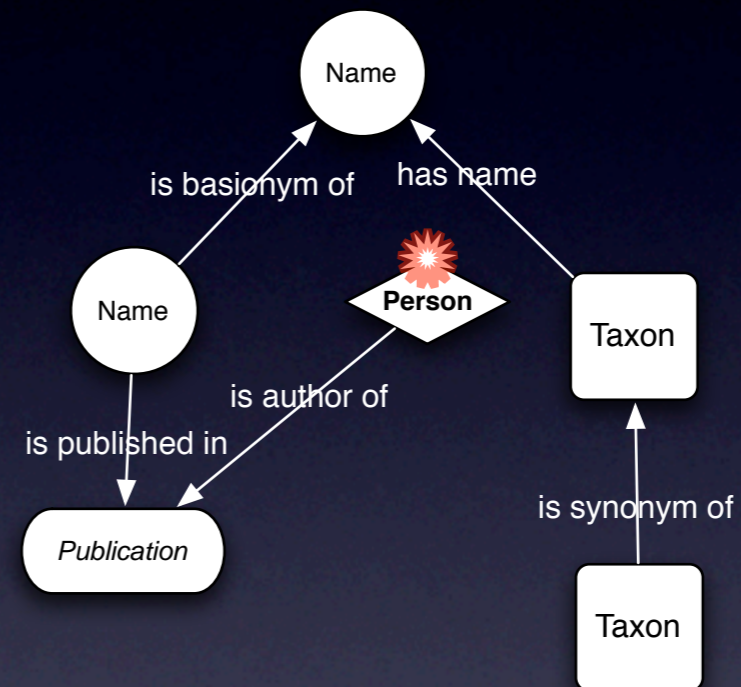
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



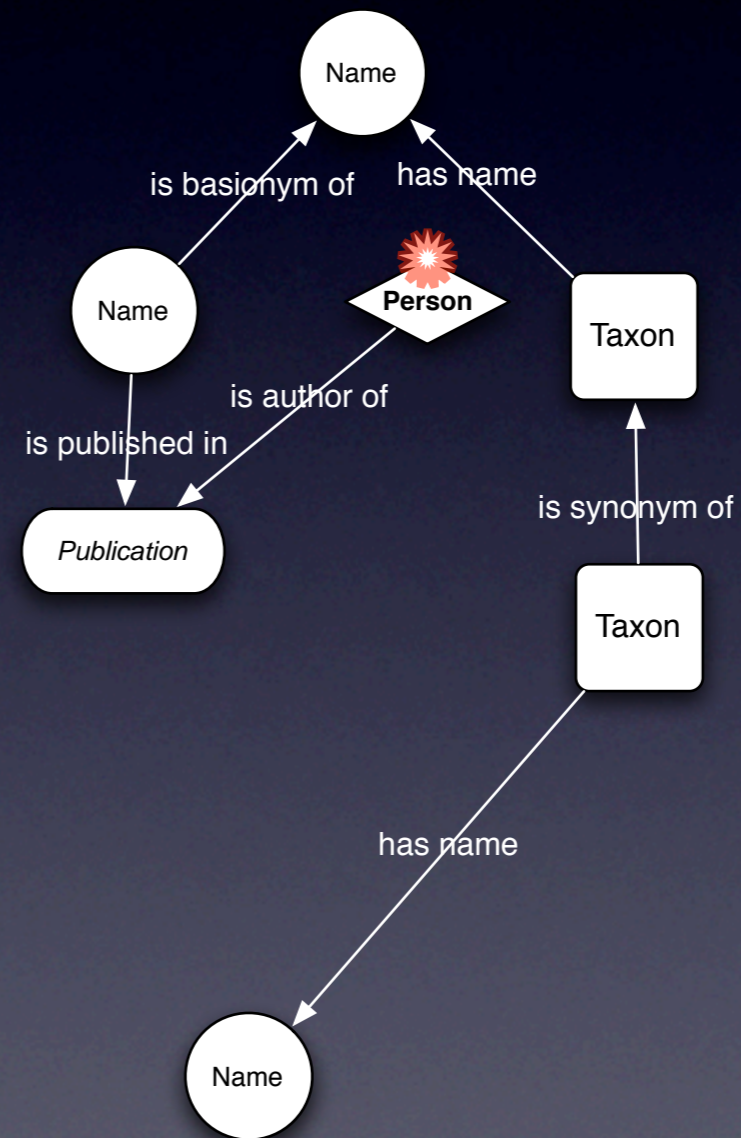
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



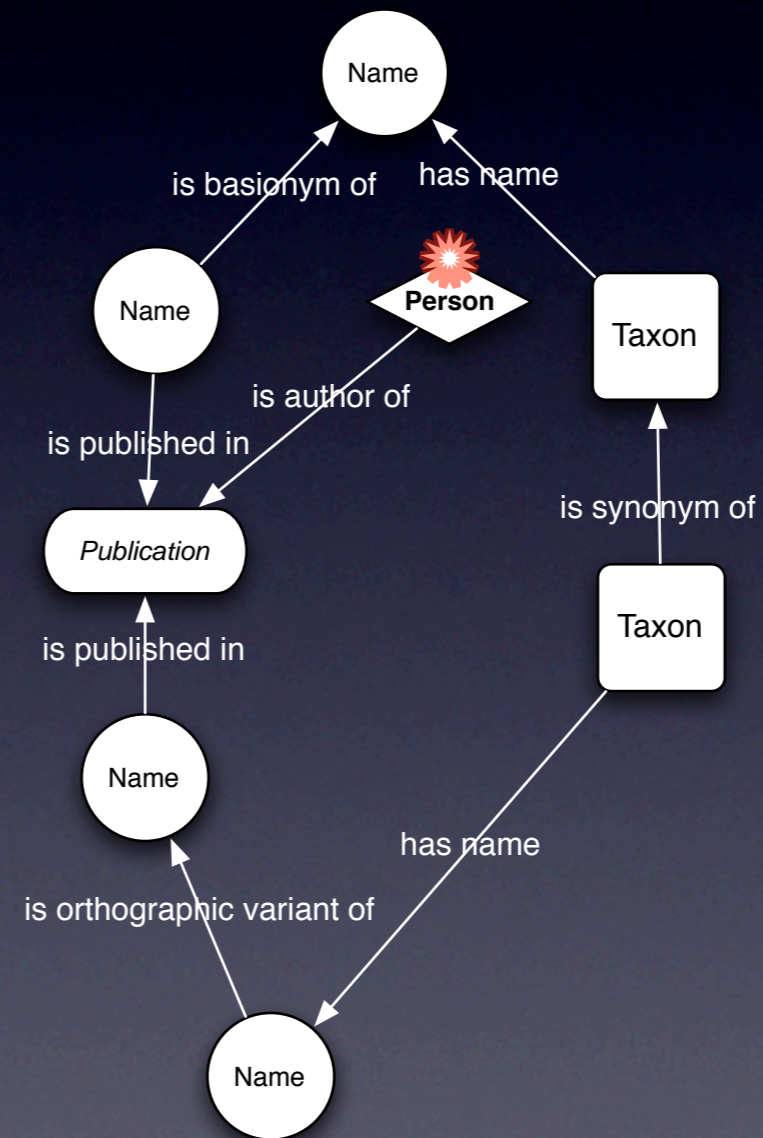
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



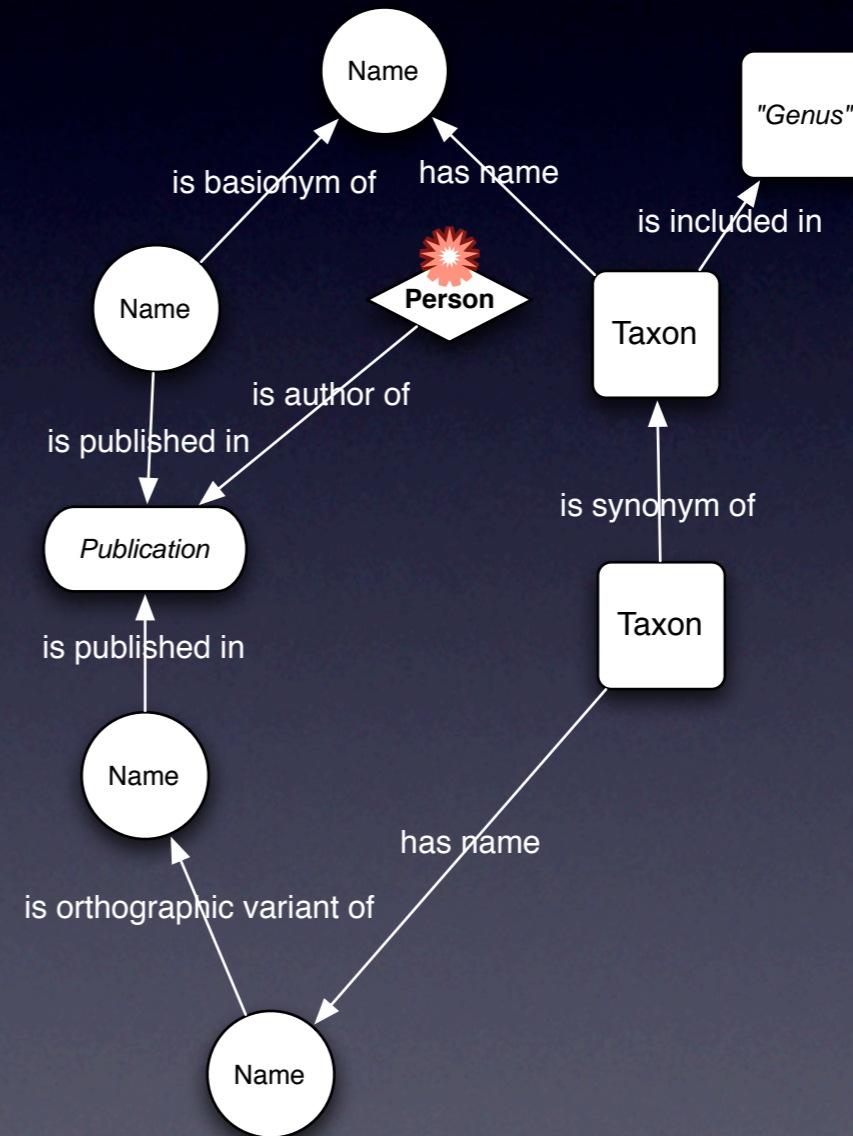
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



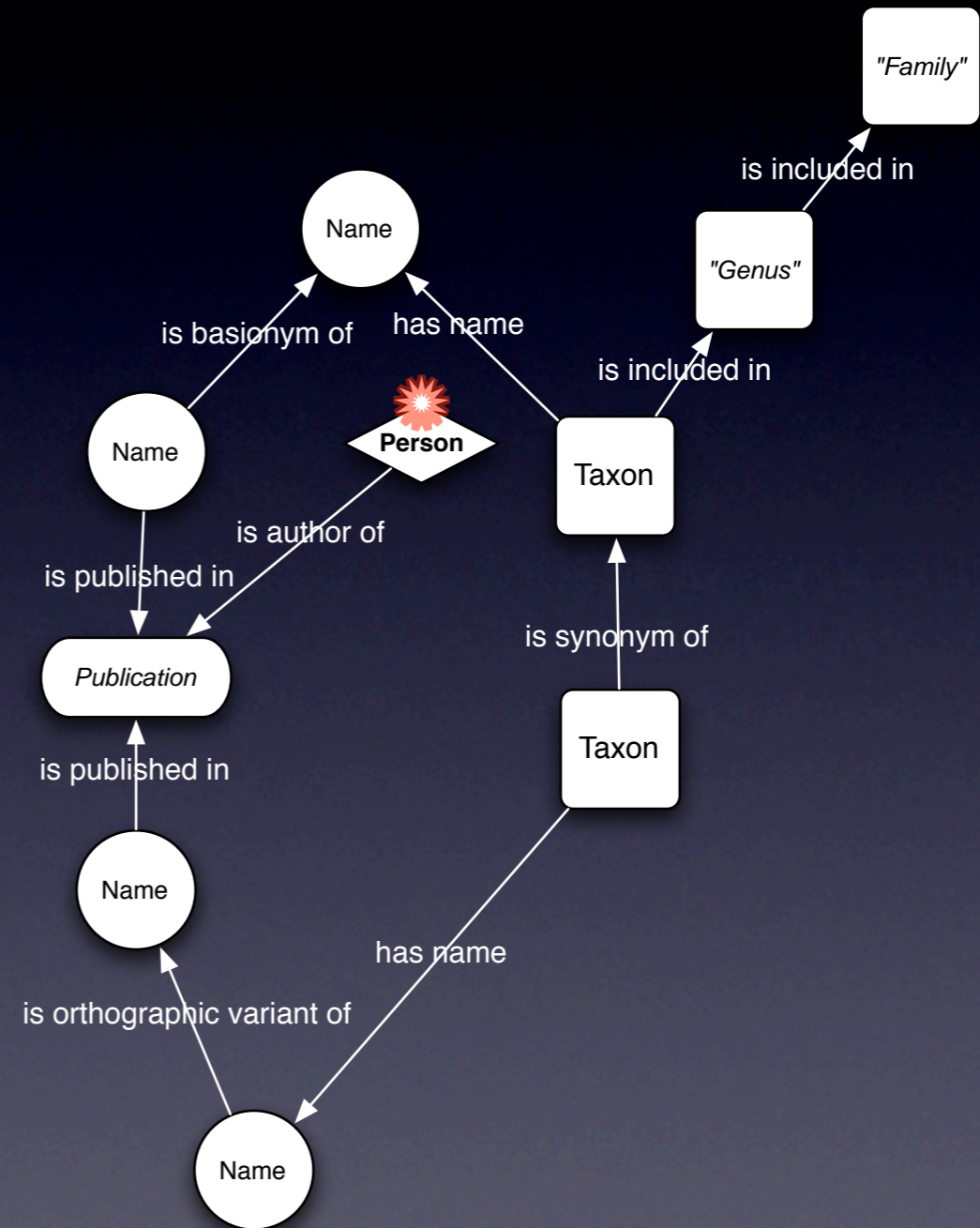
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



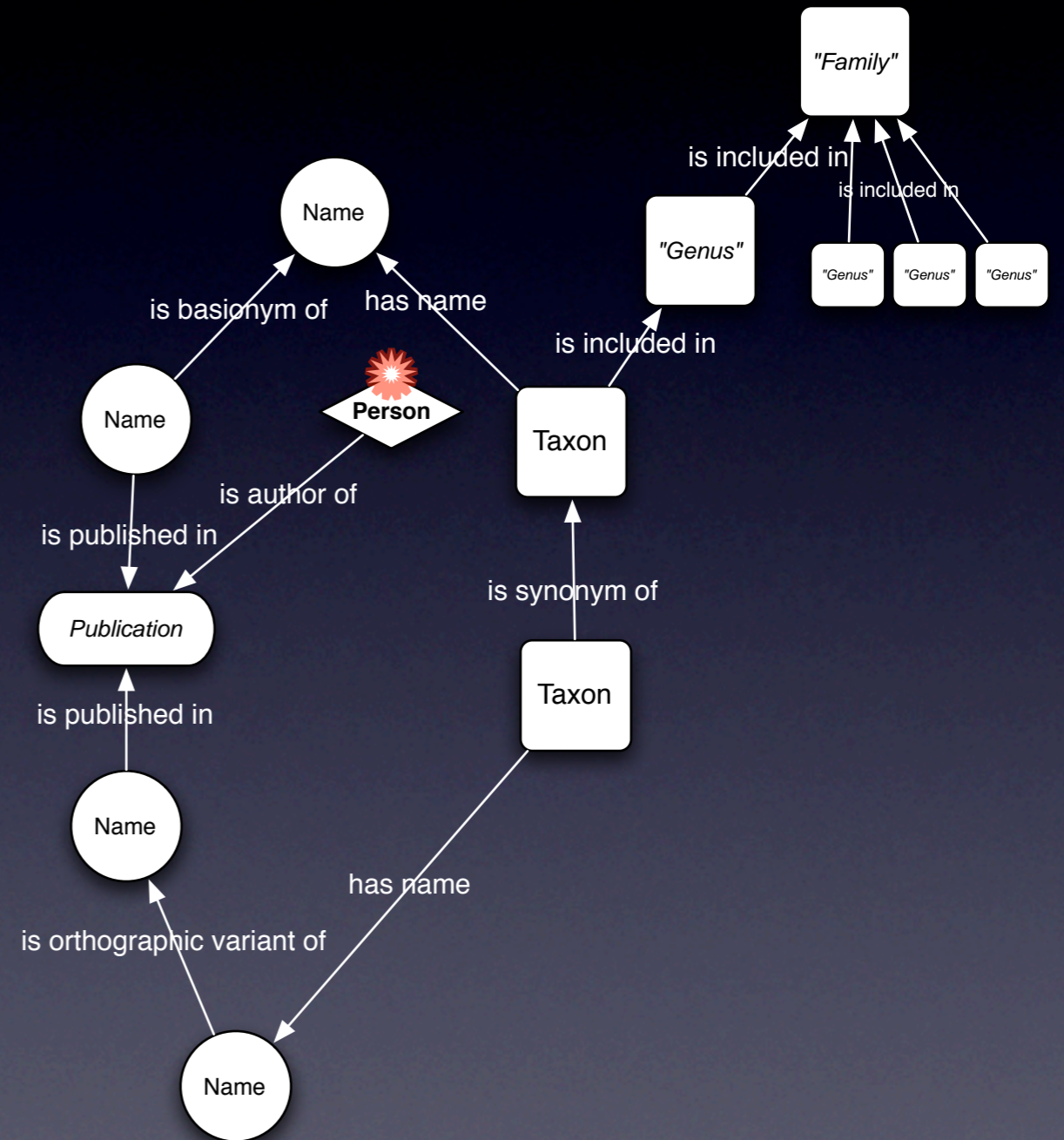
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



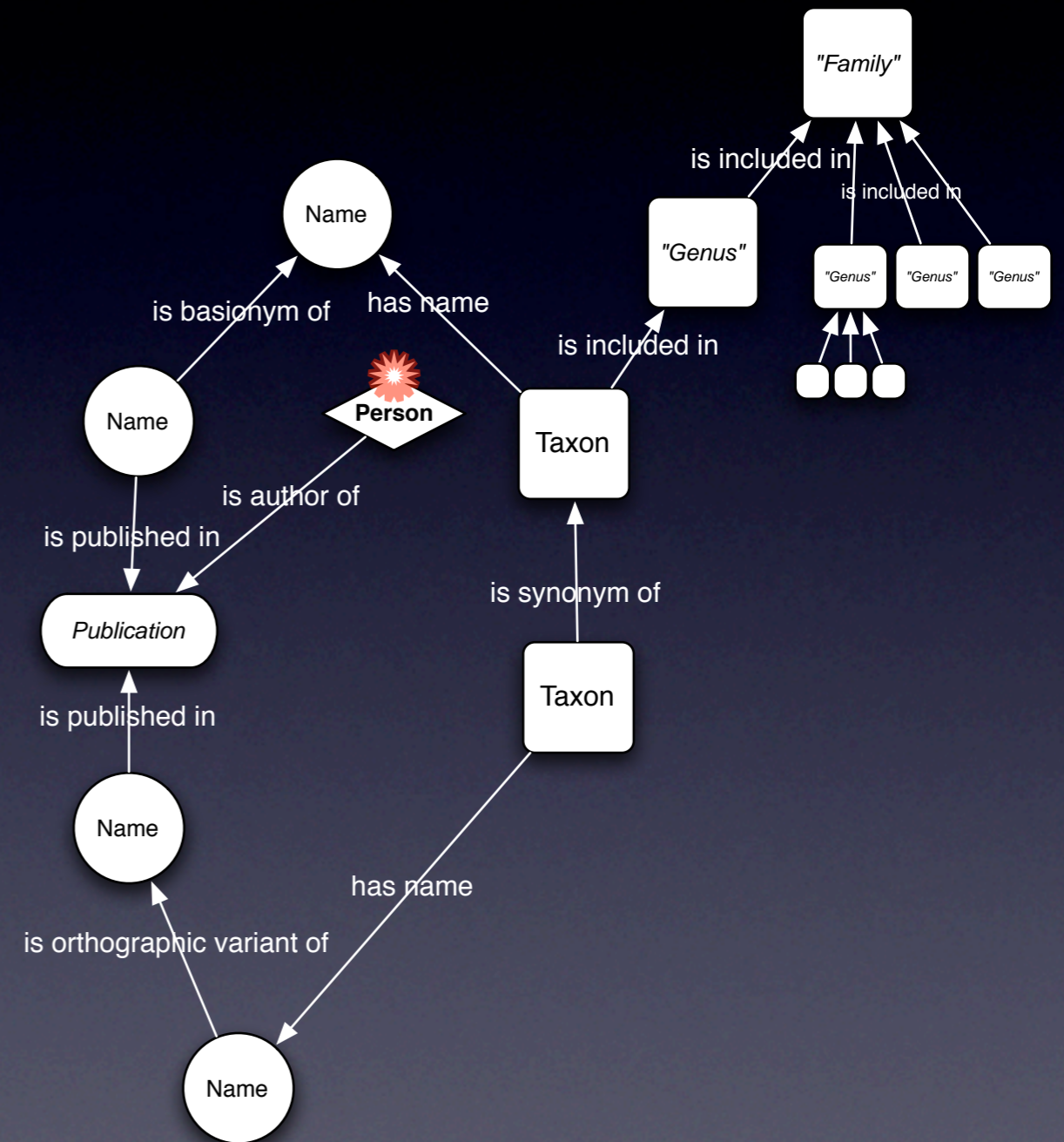
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



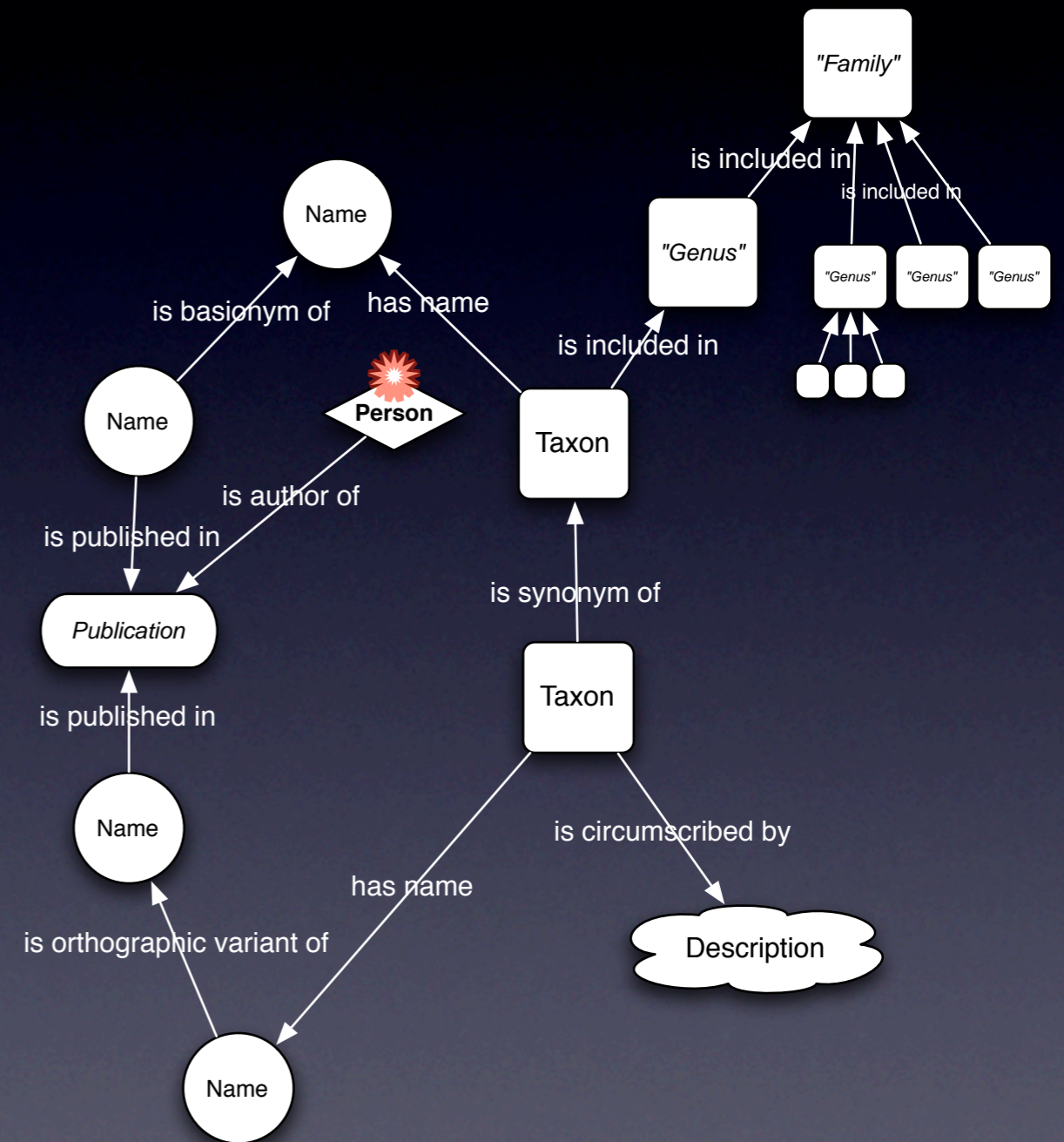
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



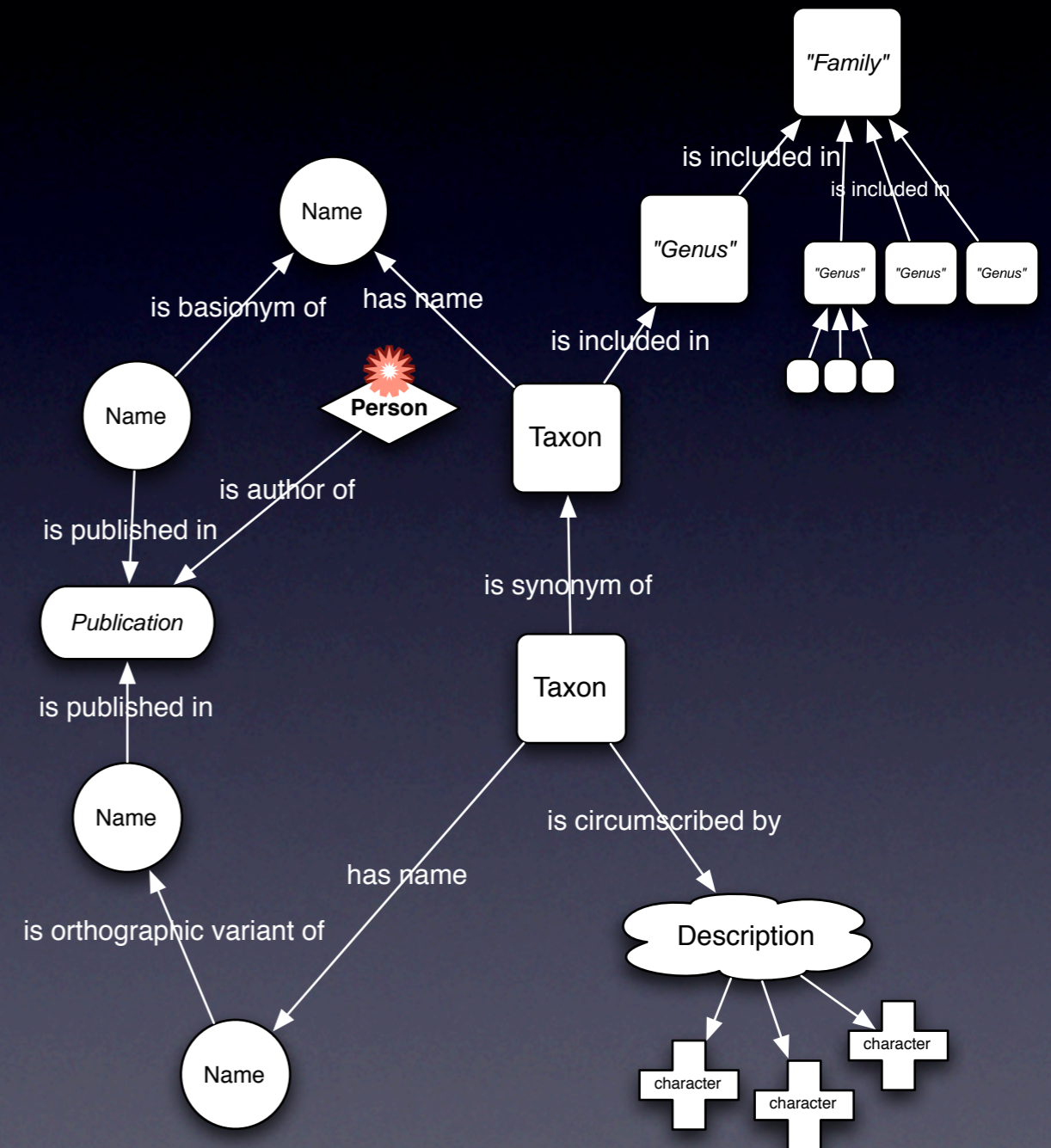
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



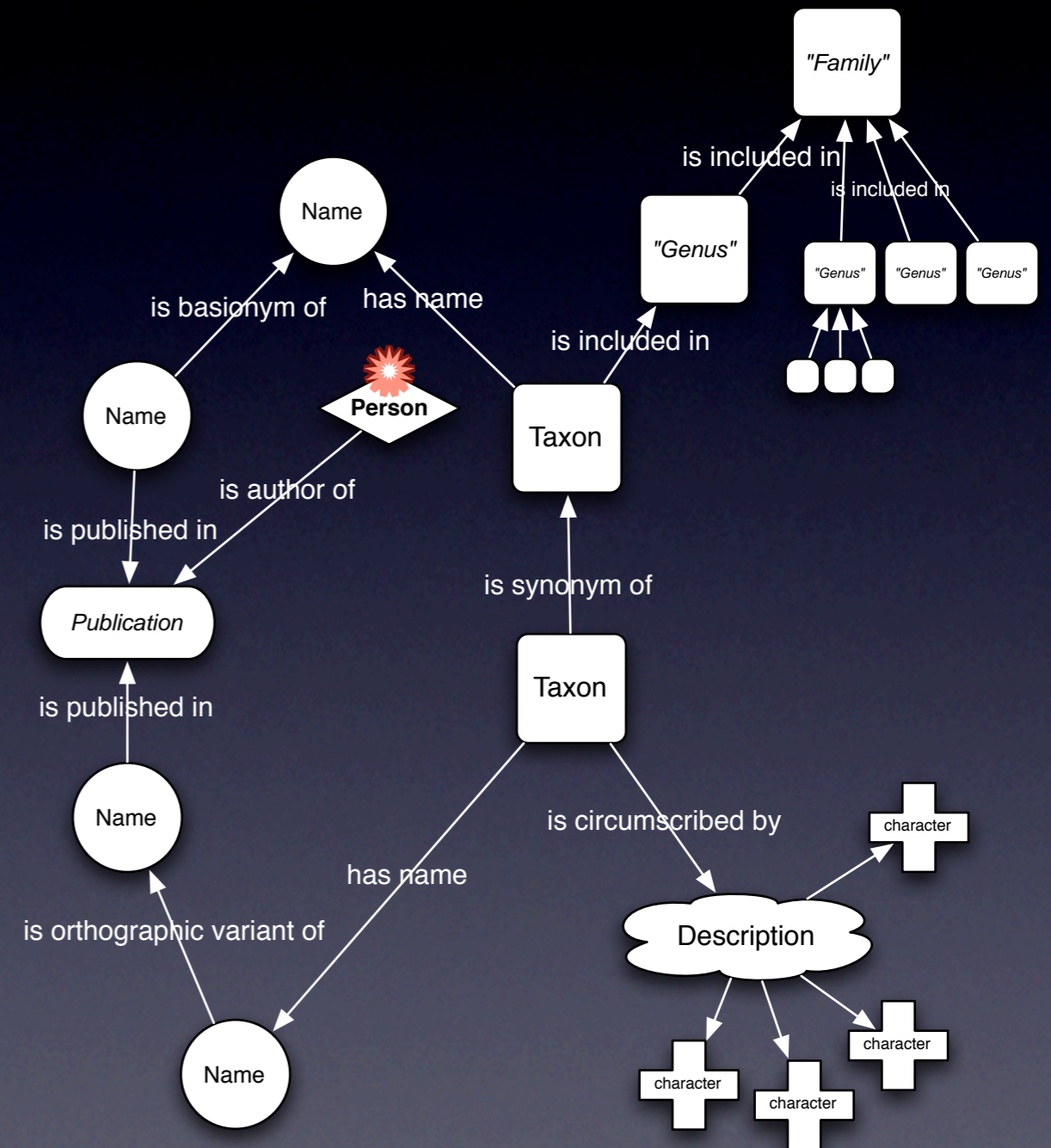
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



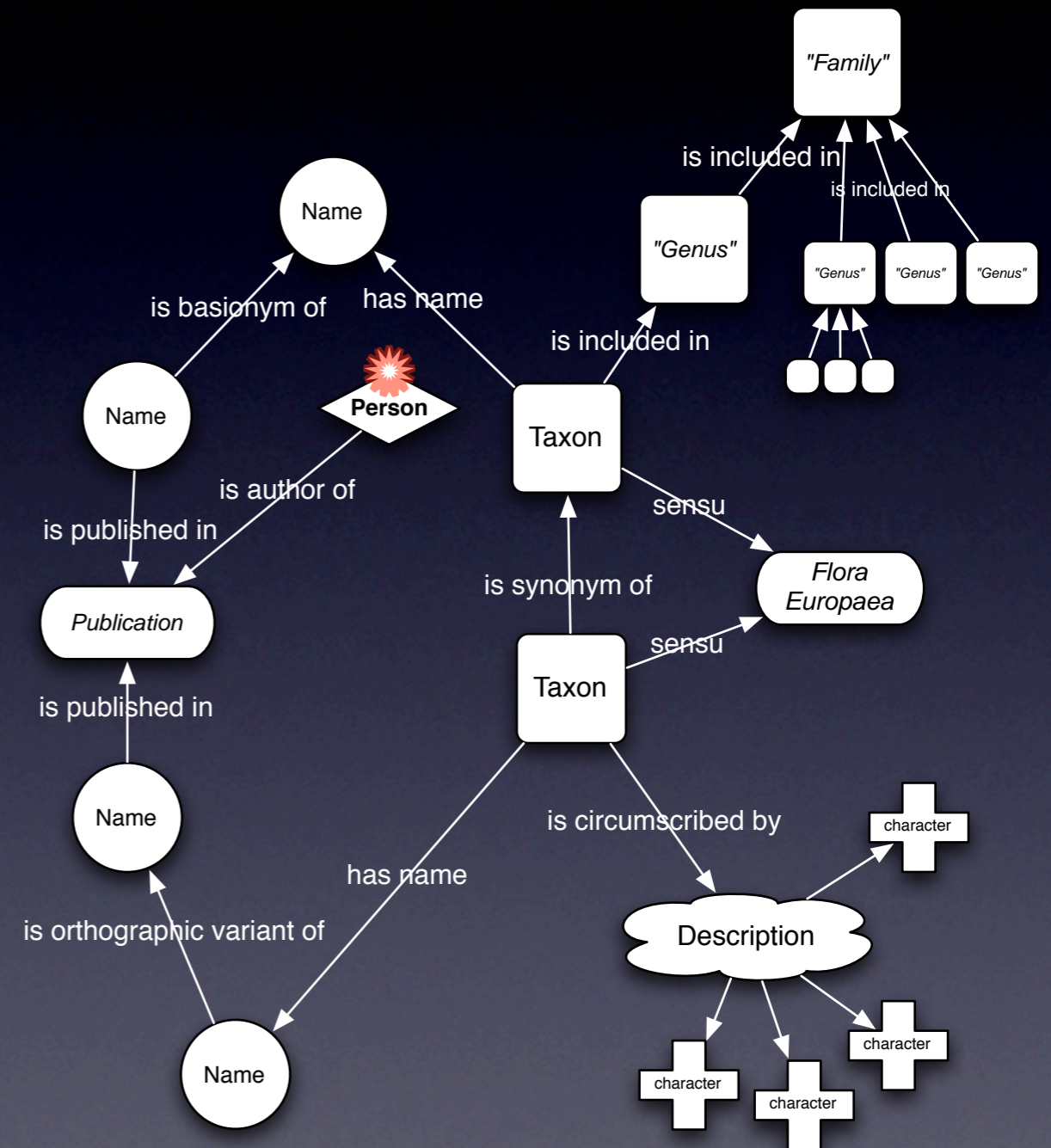
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



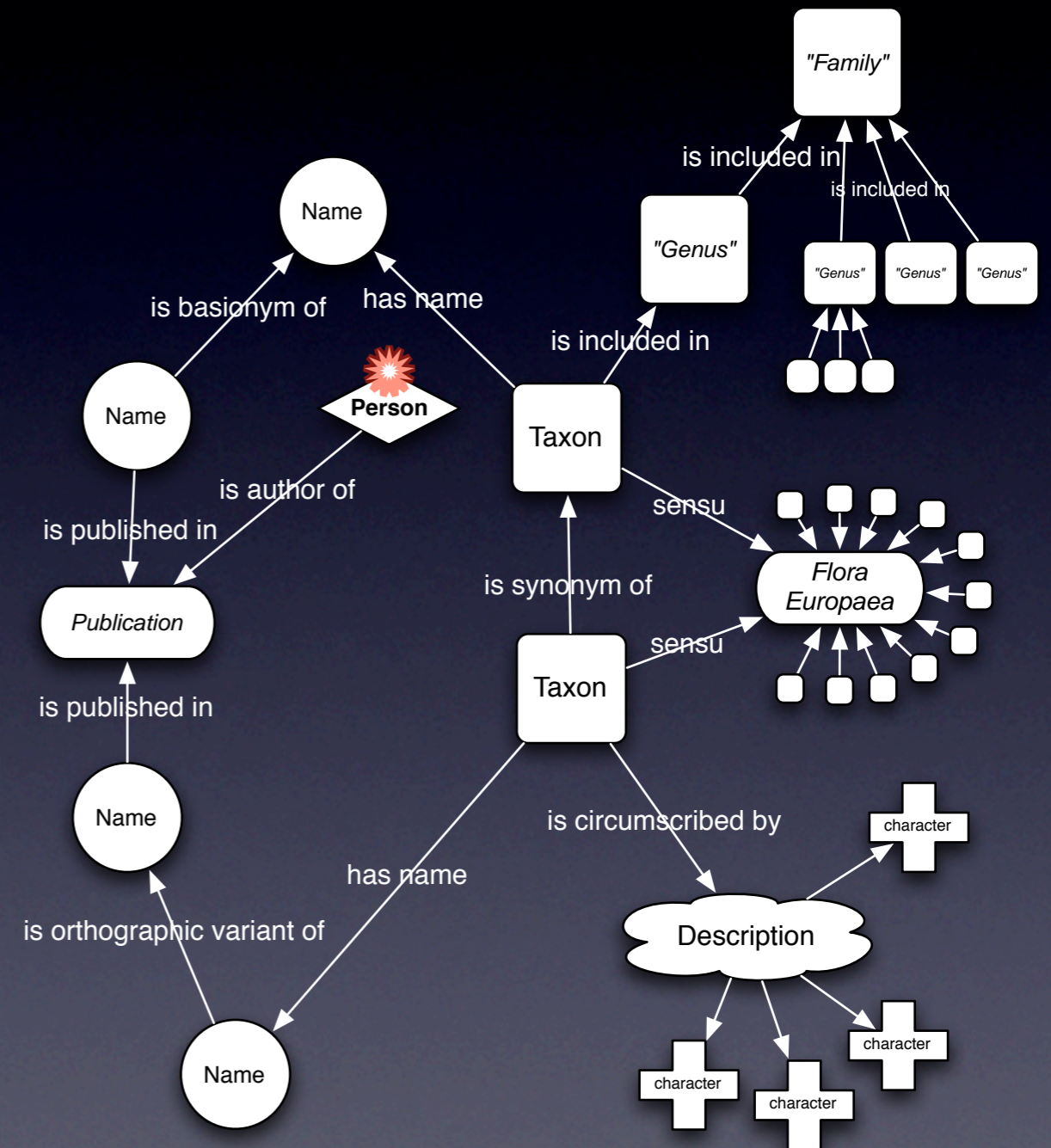
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



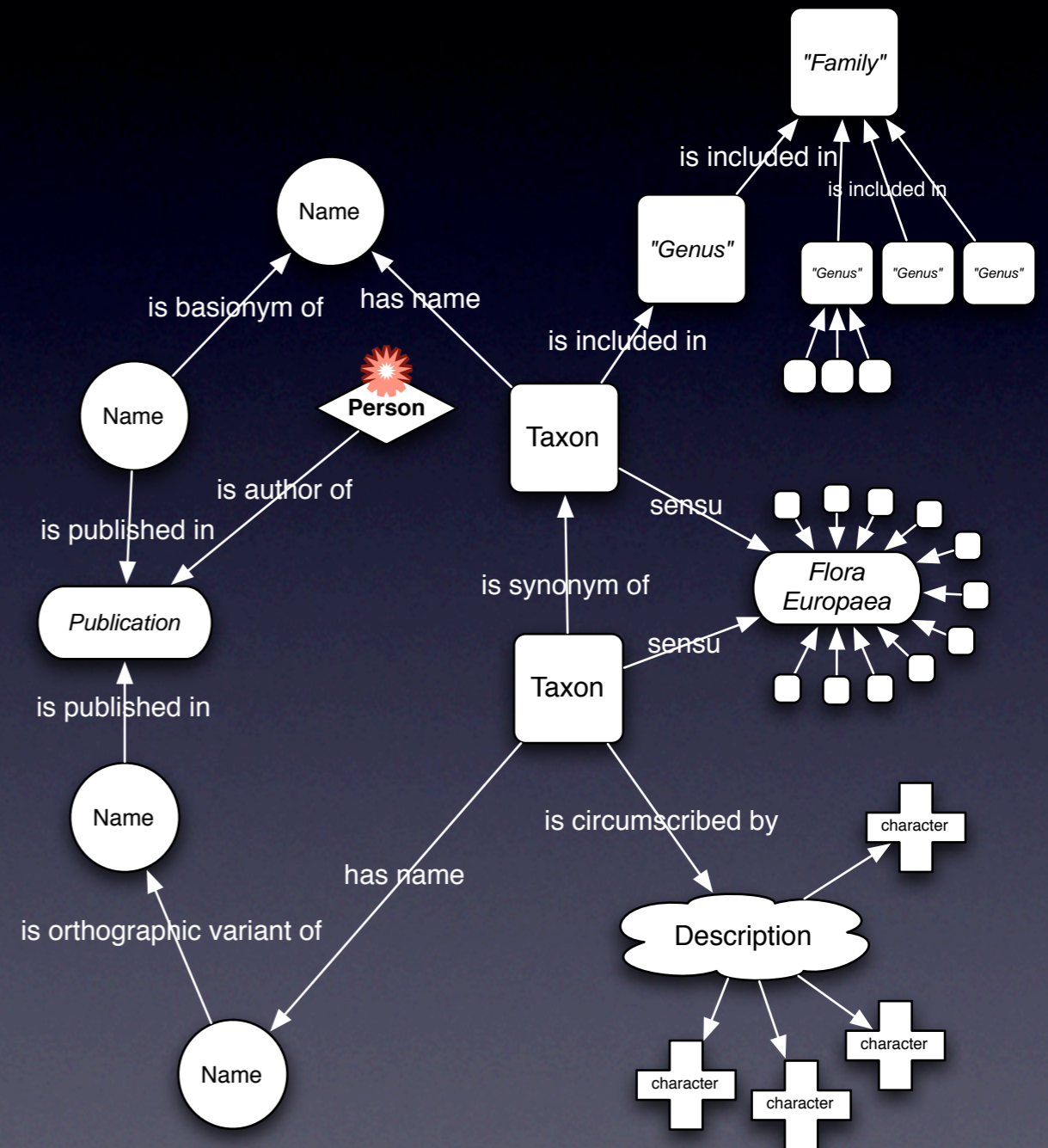
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?



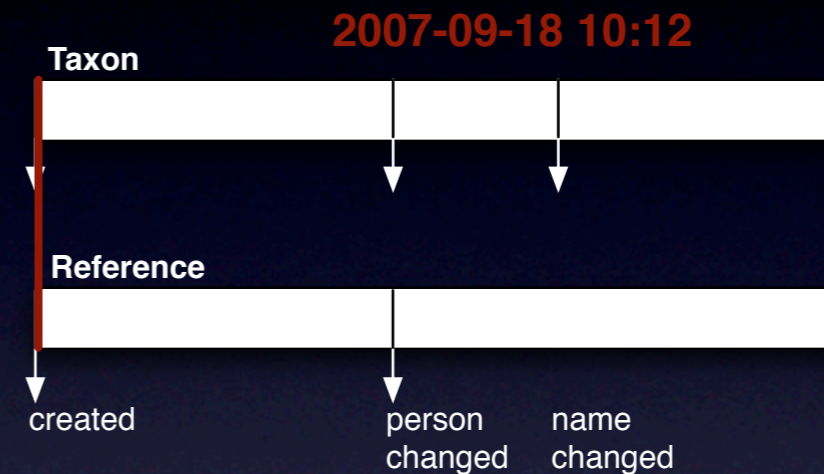
# Versioning Complex Objects

- Object graph is huge
  - potentially connecting all objects in a datastore!
- What is a version?
- Think about complex object boundaries
  - Similar to named graphs in RDF?

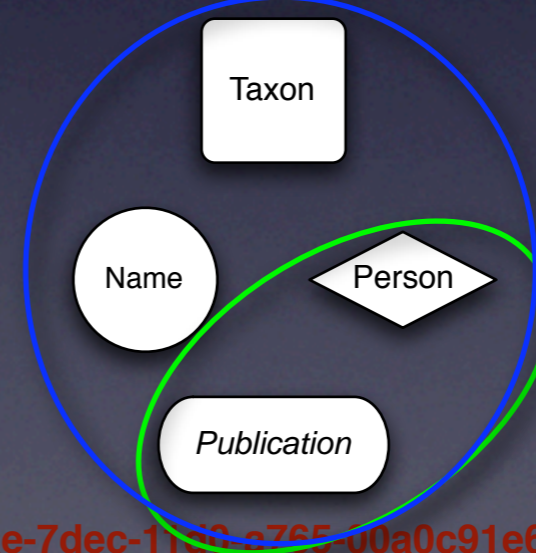


# Versioning In CDM

- Versionable base class for persistent classes
- Time based approach
  - every object has a unique identity, but different valid versions in a period of time
  - stable reference = *UUID + timestamp*
  - normalised time = time of editing
  - recovers entire datasets



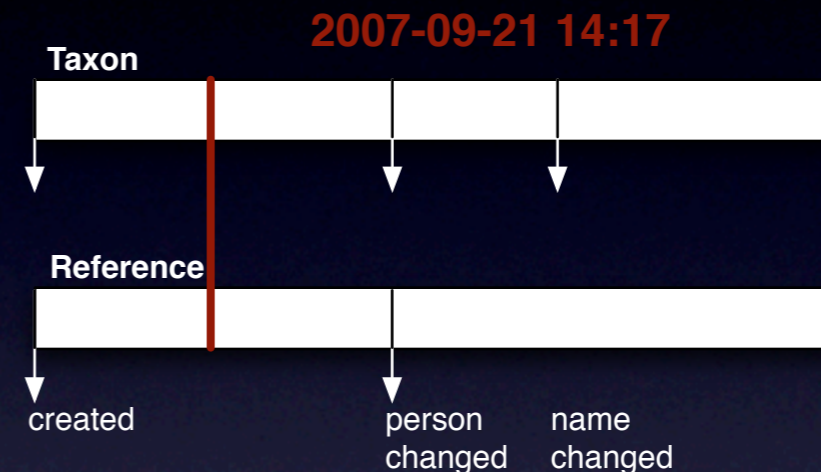
f81d4fae-7dec-11d0-a766-01c56a92a10e3 - 2007-09-18 10:12



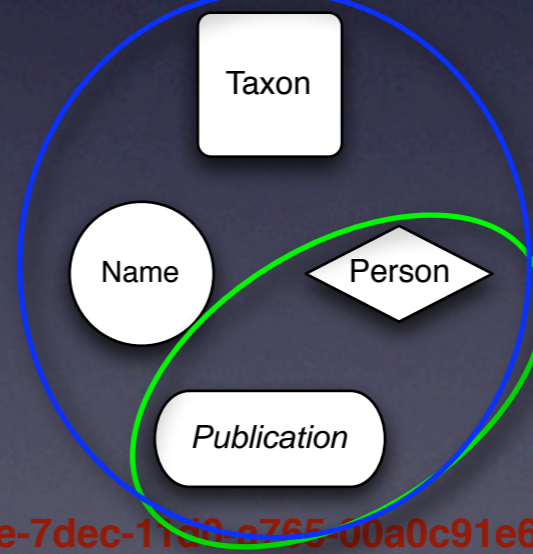
f81d4fae-7dec-11d0-a766-00a0c91e6bf6 - 2007-09-18 10:12

# Versioning In CDM

- Versionable base class for persistent classes
- Time based approach
  - every object has a unique identity, but different valid versions in a period of time
  - stable reference = *UUID + timestamp*
  - normalised time = time of editing
  - recovers entire datasets



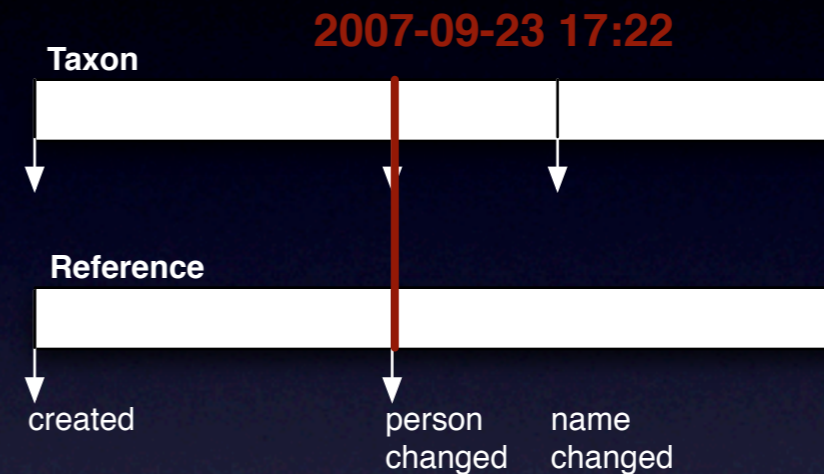
f81d4fae-7dec-11d0-a766-01c56a92a10e3 - 2007-09-18 10:12



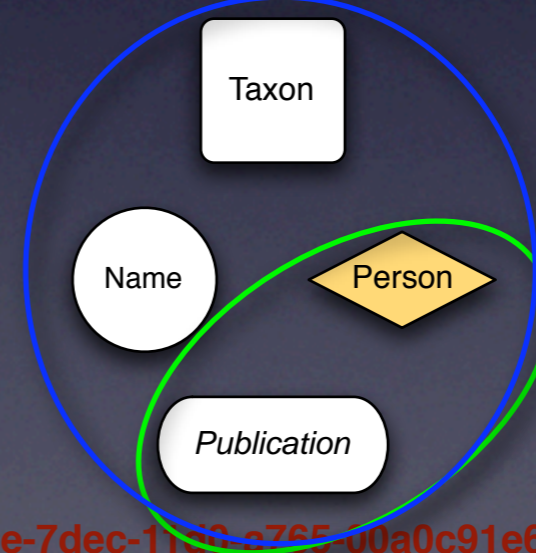
f81d4fae-7dec-11d0-a766-00a0c91e6bf6 - 2007-09-18 10:12

# Versioning In CDM

- Versionable base class for persistent classes
- Time based approach
  - every object has a unique identity, but different valid versions in a period of time
  - stable reference = *UUID + timestamp*
  - normalised time = time of editing
  - recovers entire datasets



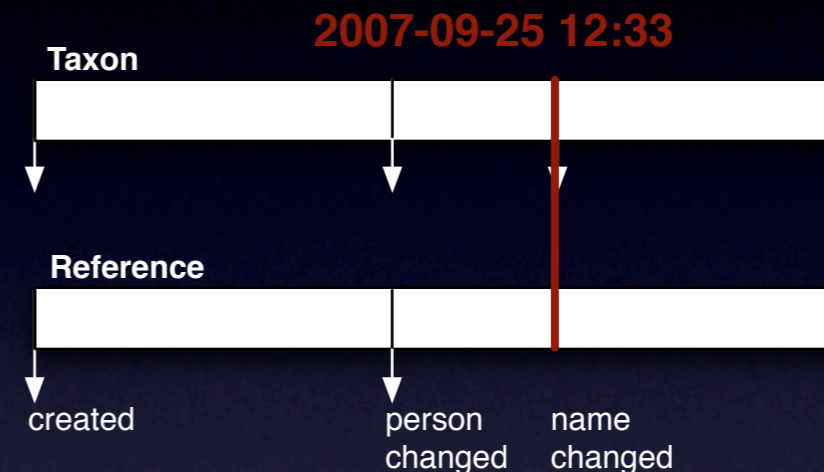
f81d4fae-7dec-11d0-a766-01c56a92a10e3 - 2007-09-23 17:22



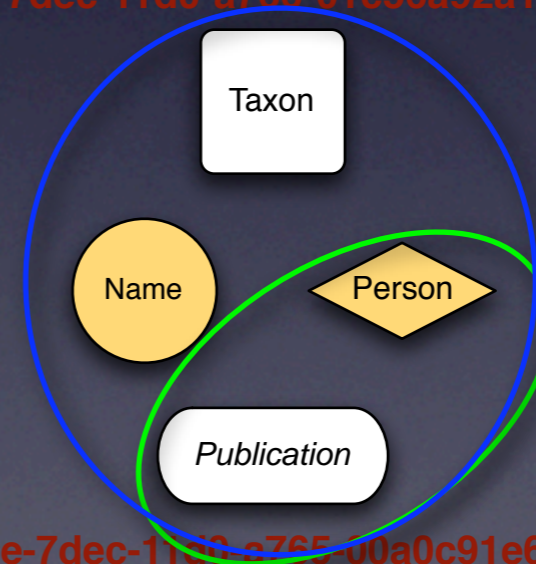
f81d4fae-7dec-11d0-a765-00a0c91e6bf6 - 2007-09-23 17:22

# Versioning In CDM

- Versionable base class for persistent classes
- Time based approach
  - every object has a unique identity, but different valid versions in a period of time
  - stable reference = *UUID + timestamp*
  - normalised time = time of editing
  - recovers entire datasets



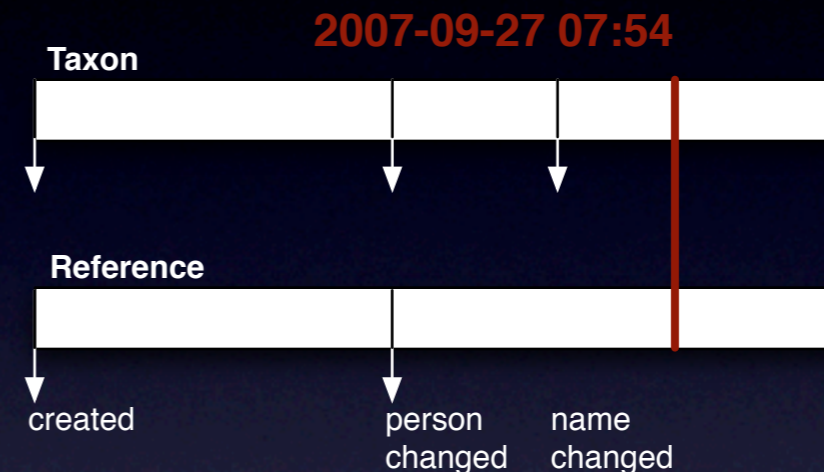
f81d4fae-7dec-11d0-a766-01c56a92a10e3 - 2007-09-25 12:33



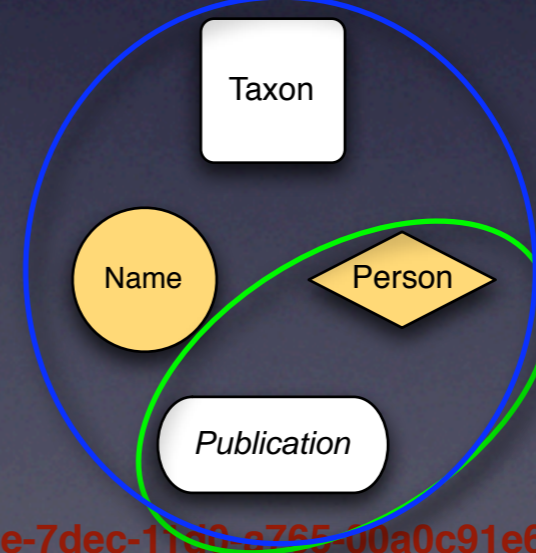
f81d4fae-7dec-11d0-a765-00a0c91e6bf6 - 2007-09-23 17:22

# Versioning In CDM

- Versionable base class for persistent classes
- Time based approach
  - every object has a unique identity, but different valid versions in a period of time
  - stable reference = *UUID + timestamp*
  - normalised time = time of editing
  - recovers entire datasets



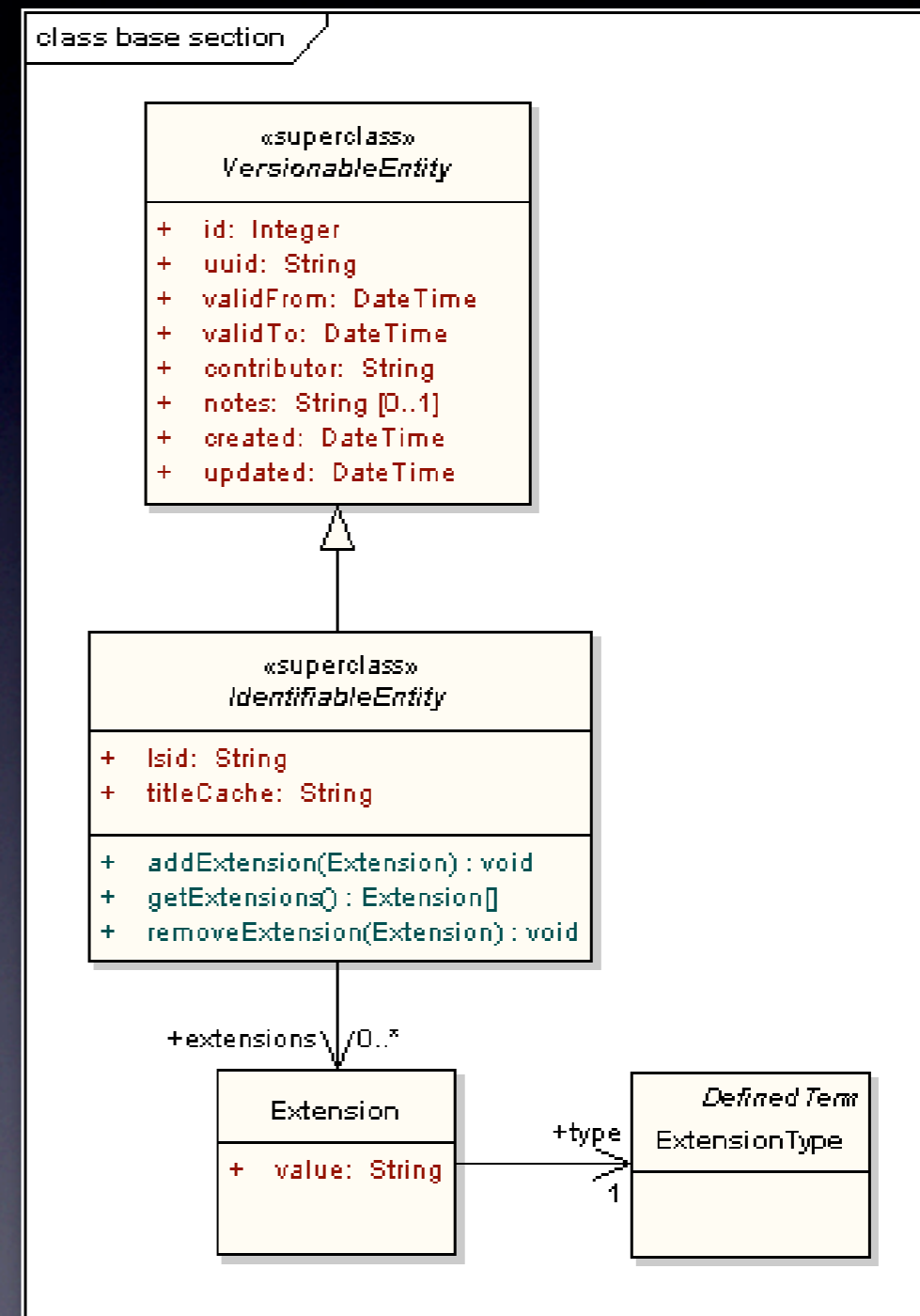
f81d4fae-7dec-11d0-a766-01c56a92a10e3 - 2007-09-25 12:33



f81d4fae-7dec-11d0-a766-00a0c91e6bf6 - 2007-09-23 17:22

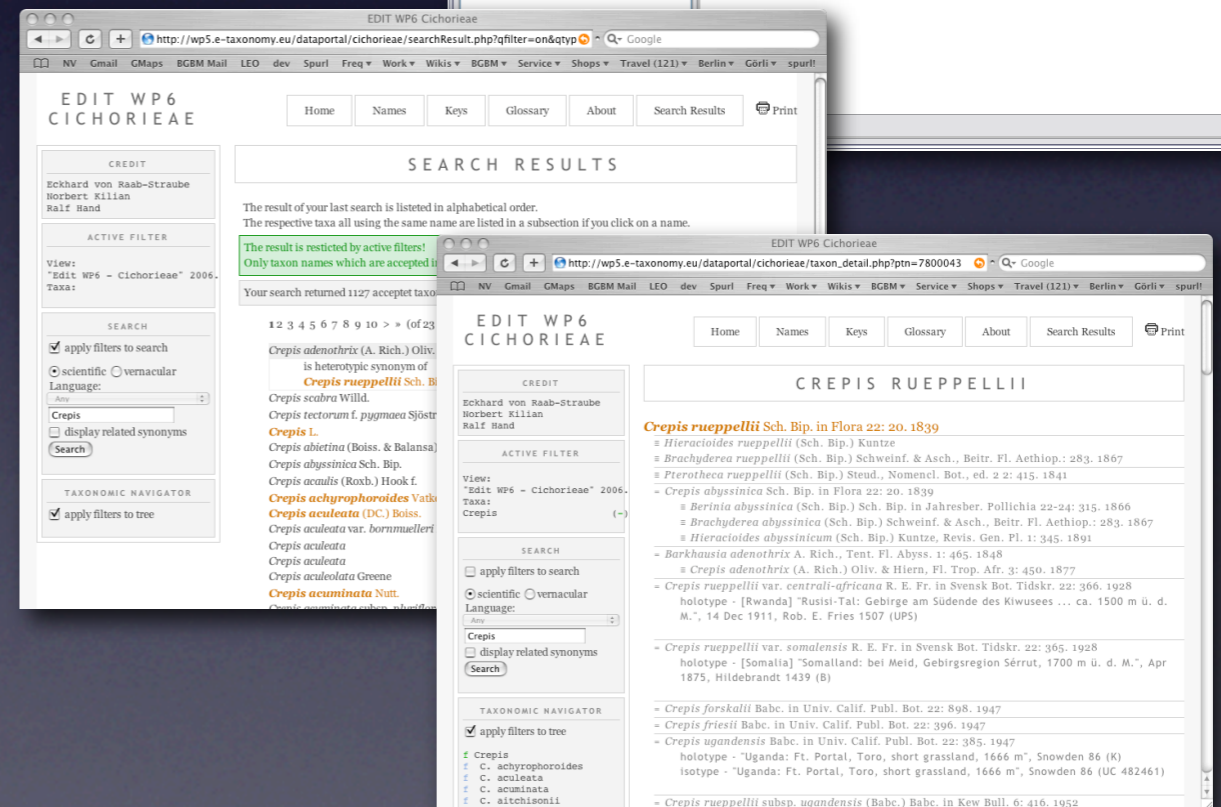
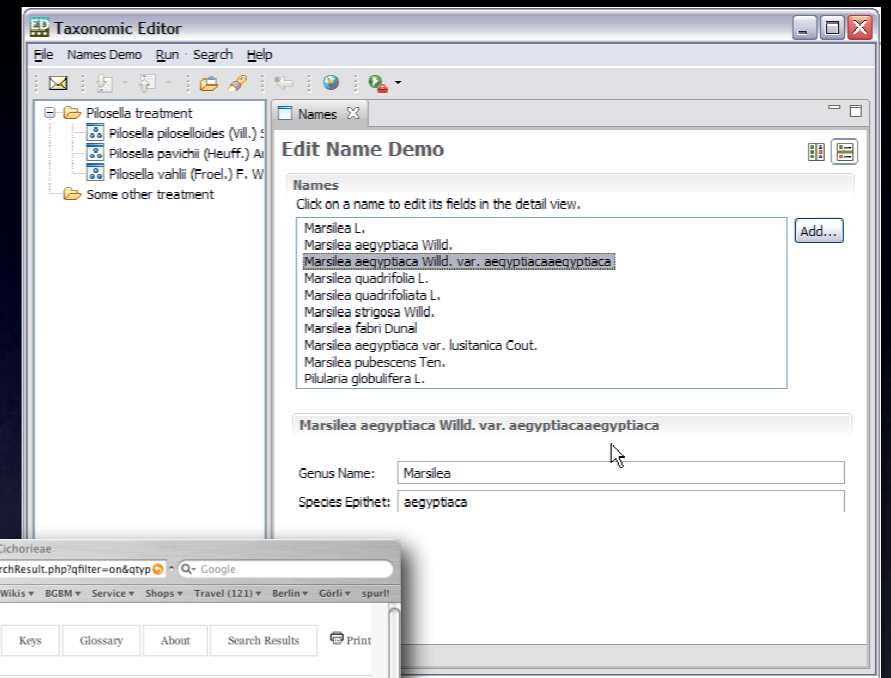
# Base Classes

- Versionable
  - UUID
  - valid period
  - contributor string
- Identifiable
  - LSID
  - title cache using default strategy
  - extension class e.g. xml/rdf fragments



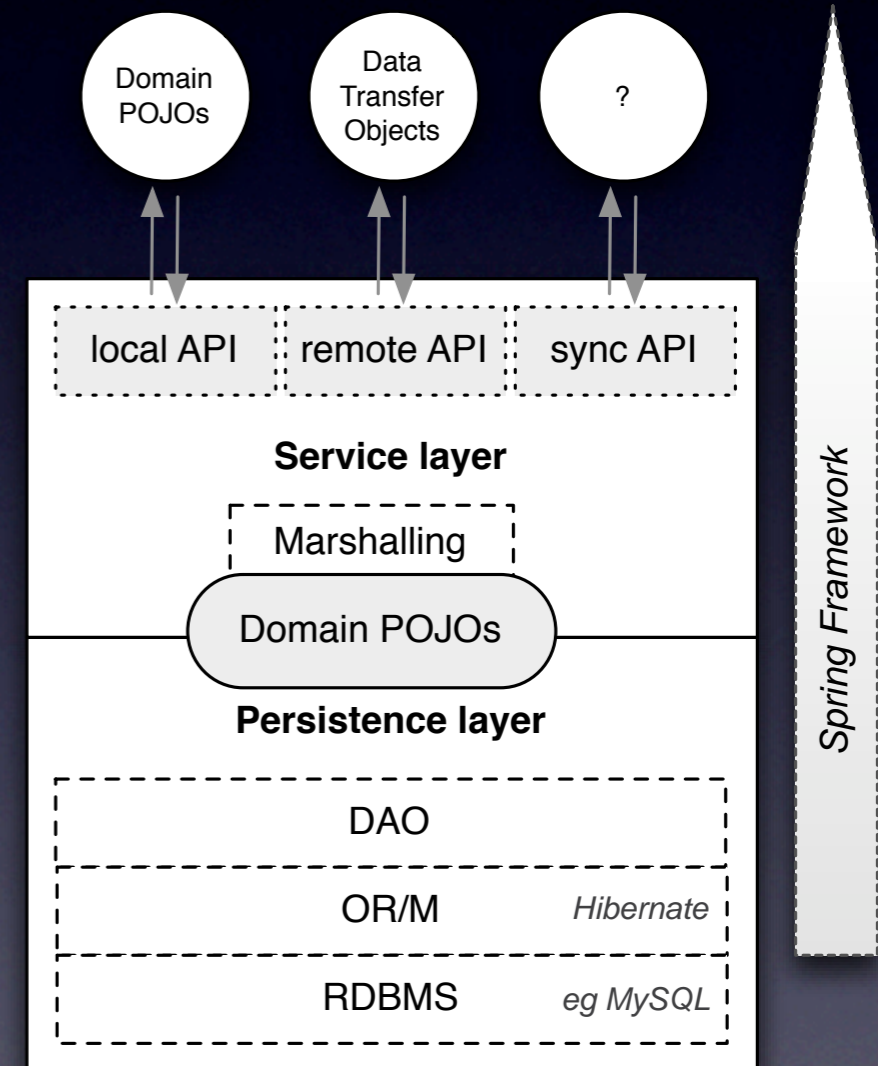
# Applications for the CDM

- Java library
  - w/ persistency
- Community Store
  - w/ services & sync API
- Taxonomic editor
  - Eclipse RCP
- Webportal
  - Drupal module
- TDWG Converter



# CDM Java Library

- Persistent domain model
  - Plain Old Java Objects
  - Hibernate & Spring framework
  - XML marshalling
  - logic & methods
    - validation, matching, ...
- Flexible use
  - several APIs
  - chose backend, incl. embedded



# Community CDM Store

- Central datastore for a community of users
- Uses CDM library w/ versioning
- Provides different APIs to interact with
  - REST read operations for portals
    - search/get for all identifiable classes. Specialised operations as required
  - RSS search feeds for subscriptions
  - Sync API, likely based on SyncML
  - OAI-PMH for aggregation
  - LSID resolver exposing objects in RDF as TDWG LSID voc

# TDWG Converter

- Converts CDM to/from TDWG LSID RDF
  - as webservice and Java source code
- No round-tripping
  - data loss from TDWG to CDM
  - but serialised RDF triples could be stored as extensions
- Can you safely ignore parts of the RDF data?
  - not understood triples can change semantics
  - like a letter with a foreign language at some paragraphs

# Conclusion

- CDM library & store freely available in 2008
- TDWG ontology important for defining semantics
  - external vocabulary applicabilities, e.g. languages, countries
  - unify SDD, SPM and maybe LTER/EML
  - consistent, integrated XML schema might be worth too
- Think about complex object boundaries

[cdm-edit@mnhn.fr](mailto:cdm-edit@mnhn.fr)

<http://dev.e-taxonomy.eu/wiki/CommonDataModel>