Visualization of Inferred Versioned Schemas from NoSQL Databases

Alberto Hernández Chillón
alberto.hernandez1@um.es
Cátedra SAES-UMU
University of Murcia

Diego Sevilla Ruiz
dsevilla@ditec.um.es
Faculty of Computer Science
University of Murcia

Jesús García Molina
jmolina@um.es
Faculty of Computer Science
University of Murcia

November 15, 2016

SiriusCON|16

UNIVERSIDAD DE MURCIA

About the authors

- Member of the Cátedra SAES team since 2014
- M.Sc. in Computer Science from the UM
- MDE, automatic code generation, NoSQL databases

- Associate professor at the Faculty of Computer Science of the UM
- M.Sc and Ph.D. in Computer Science from the UM
- NoSQL databases, distributed systems, testing

- Professor at the Faculty of Computer Science of the UM since 1984
- Head of the Modelum Group
- MDE, DSL, Software modernization, reverse engineering

SiriusCON 16

UNIVERSIDAD DE MURCIA

Index

SiriusCON|16

UNIVERSIDAD DE MURCIA

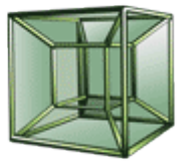
Index

SiriusCON 16

UNIVERSIDAD DE MURCIA

Introduction and context

NoSQL systems



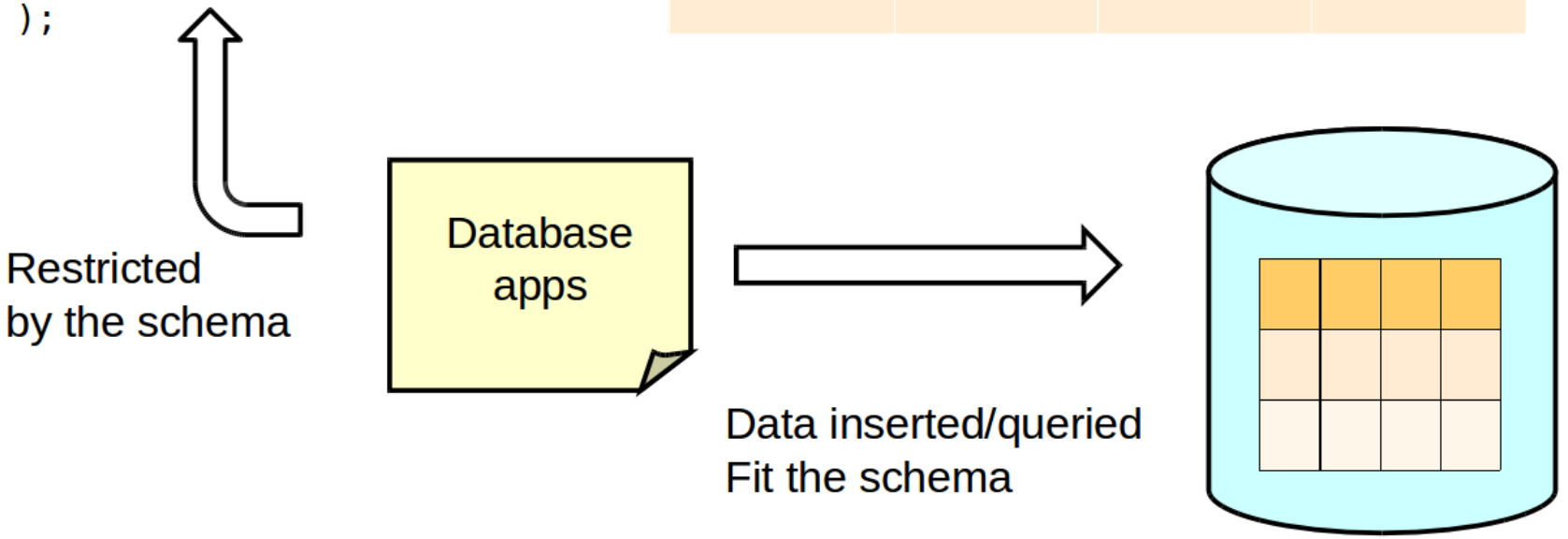
SiriusCON 16

UNIVERSIDAD DE MURCIA

Usefulness of an explicit schema

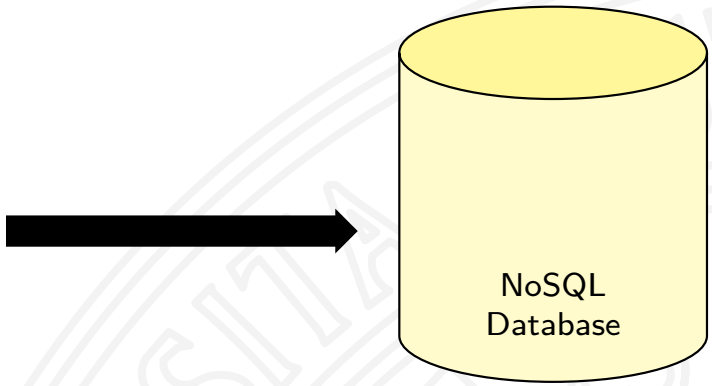
```
CREATE TABLE Persons (
  PersonID int,
  LastName varchar(255),
  FirstName varchar(255),
  Address varchar(255)
);
```

PersonID	LastName	FirstName	Address



What kind of NoSQL systems? - Schemaless

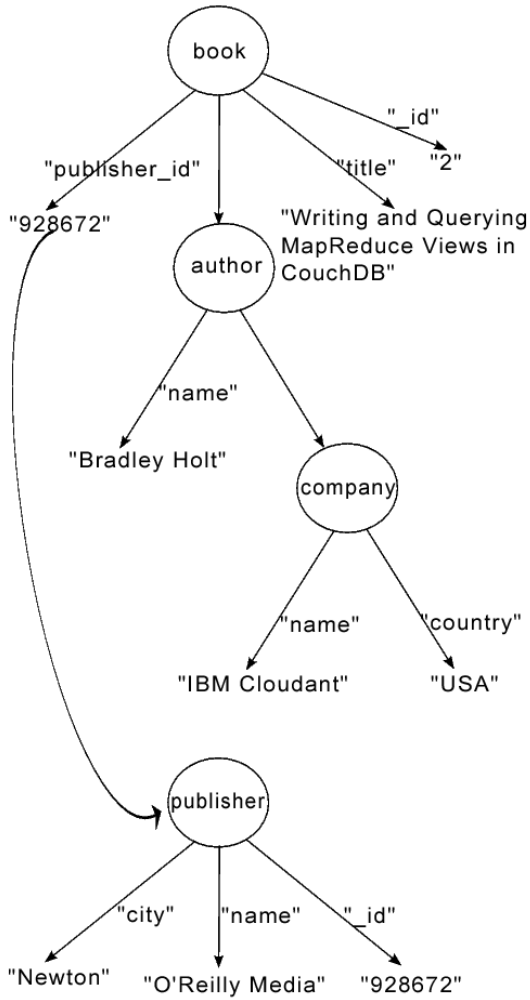
```
{
  "person_id": "123",
  "type": "Person",
  "lastName": "Rush",
  "firstName": "Christopher",
  "address": "C/Gran Via, 13, Madrid"
},
{
  "person_id": "456",
  "type": "Person",
  "lastName": "England",
  "firstName": "Wayne",
  "address": {
    "street": "Av. Pinos, 24",
    "city": "Murcia" }
},
{
  "person_id": "789",
  "type": "Person",
  "lastName": "Hoover",
  "firstName": "Quinton",
  "address": "Ronda Norte, 15, Murcia",
  "age": 35
}
```



- Data non-uniformity
- Different versions for the same data

What kind of NoSQL systems? - Aggregations

```
{
  "_id": "2",
  "type": "book",
  "title": "Writing and Querying MapReduce
       Views in CouchDB",
  "publisher_id": "928672",
  "author":
  {
    "_id": "101",
    "type": "author",
    "name": "Bradley Holt",
    "company": {
      "_id": "324",
      "type": "company",
      "name": "IBM Cloudant",
      "country": "USA"}
  },
},
{
  "_id": "928672",
  "type": "publisher",
  "name": "O'Reilly Media",
  "city": "Newton"
}
```



SiriusCON 16

UNIVERSIDAD DE MURCIA

Proposed objective

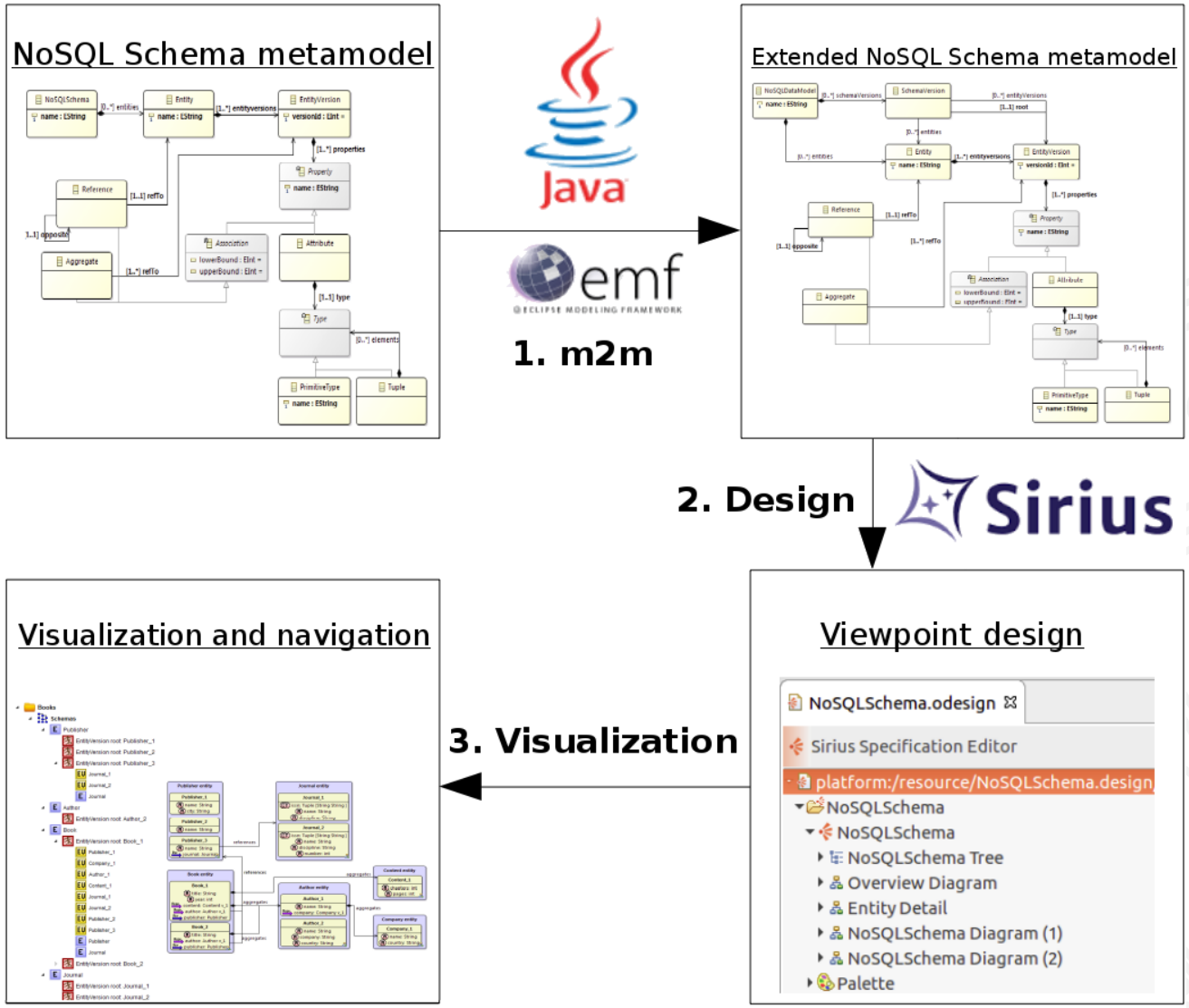
- A *DataVersity* report (2015) indicated essential functionality required in the near future:
  - **Model visualization**
  - Code generation from schemas
  - Metadata management
- Our goal is to design and implement a tool which will allow us to visualize NoSQL schemas:
  - Taking into account concepts such as versions...
  - ...having in mind the underlying inference process...
  - ...among other things

Index

SiriusCON 16

UNIVERSIDAD DE MURCIA

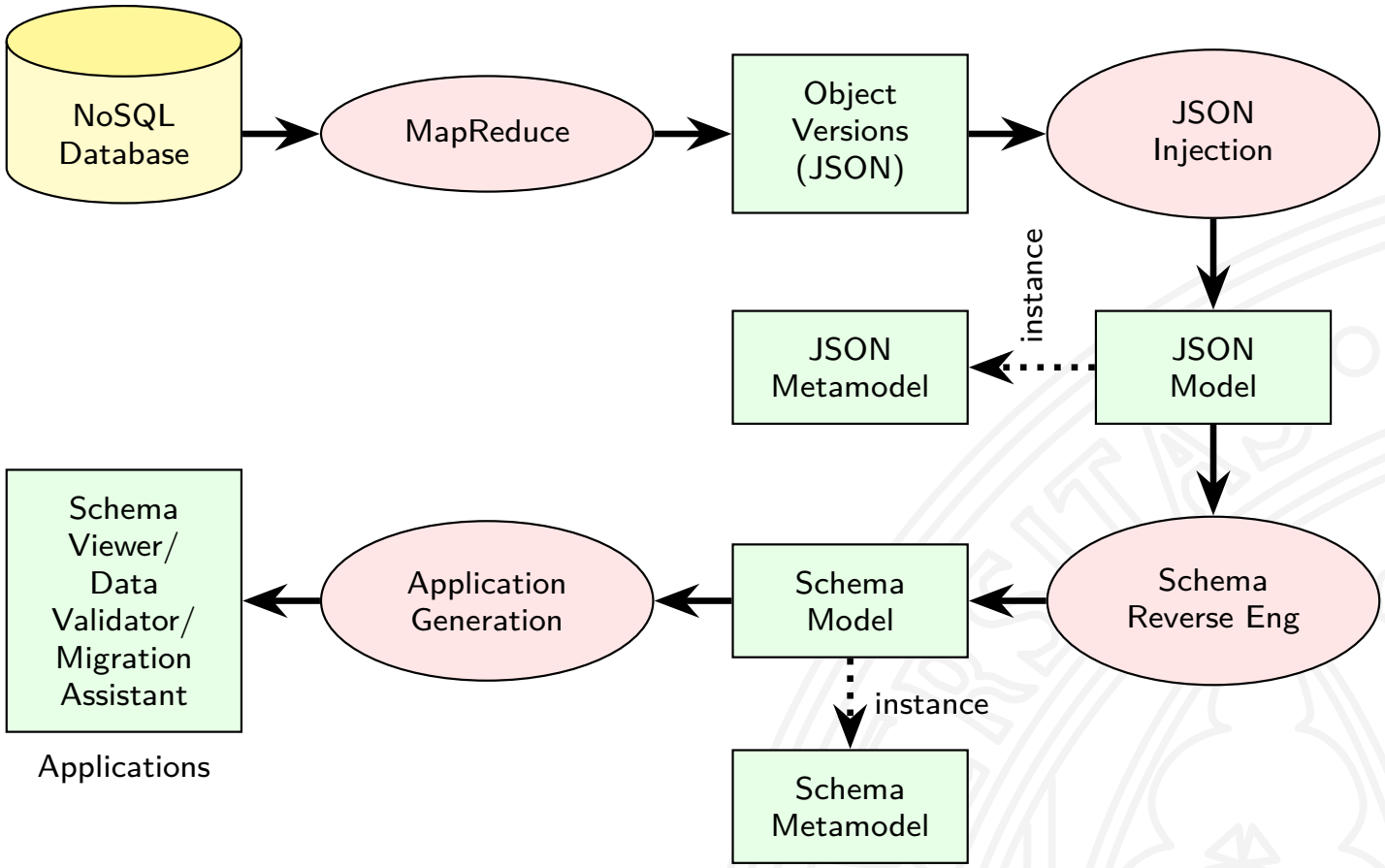
Process overview



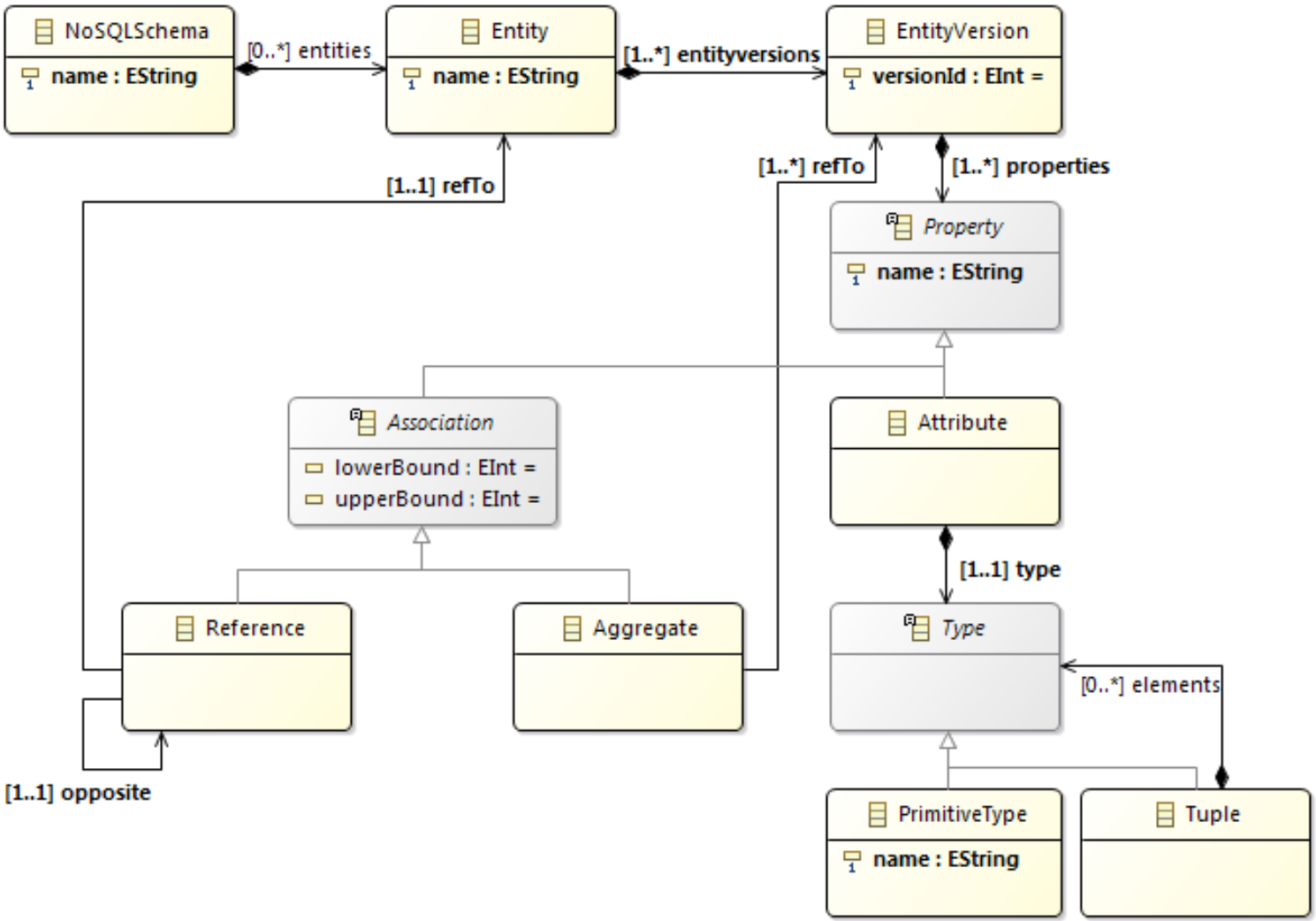
SiriusCON 16

UNIVERSIDAD DE MURCIA

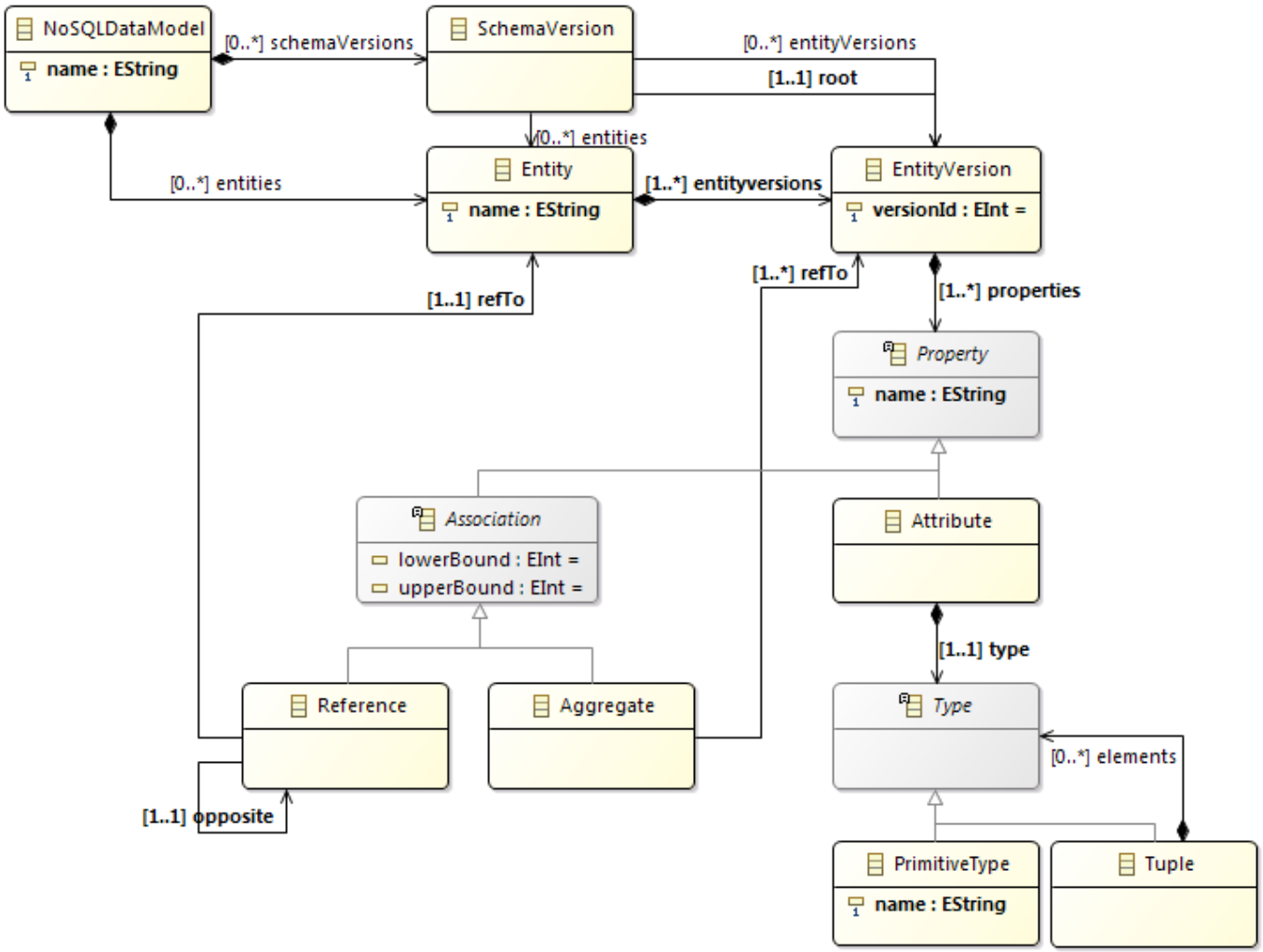
Inference process (I)



Initial metamodel



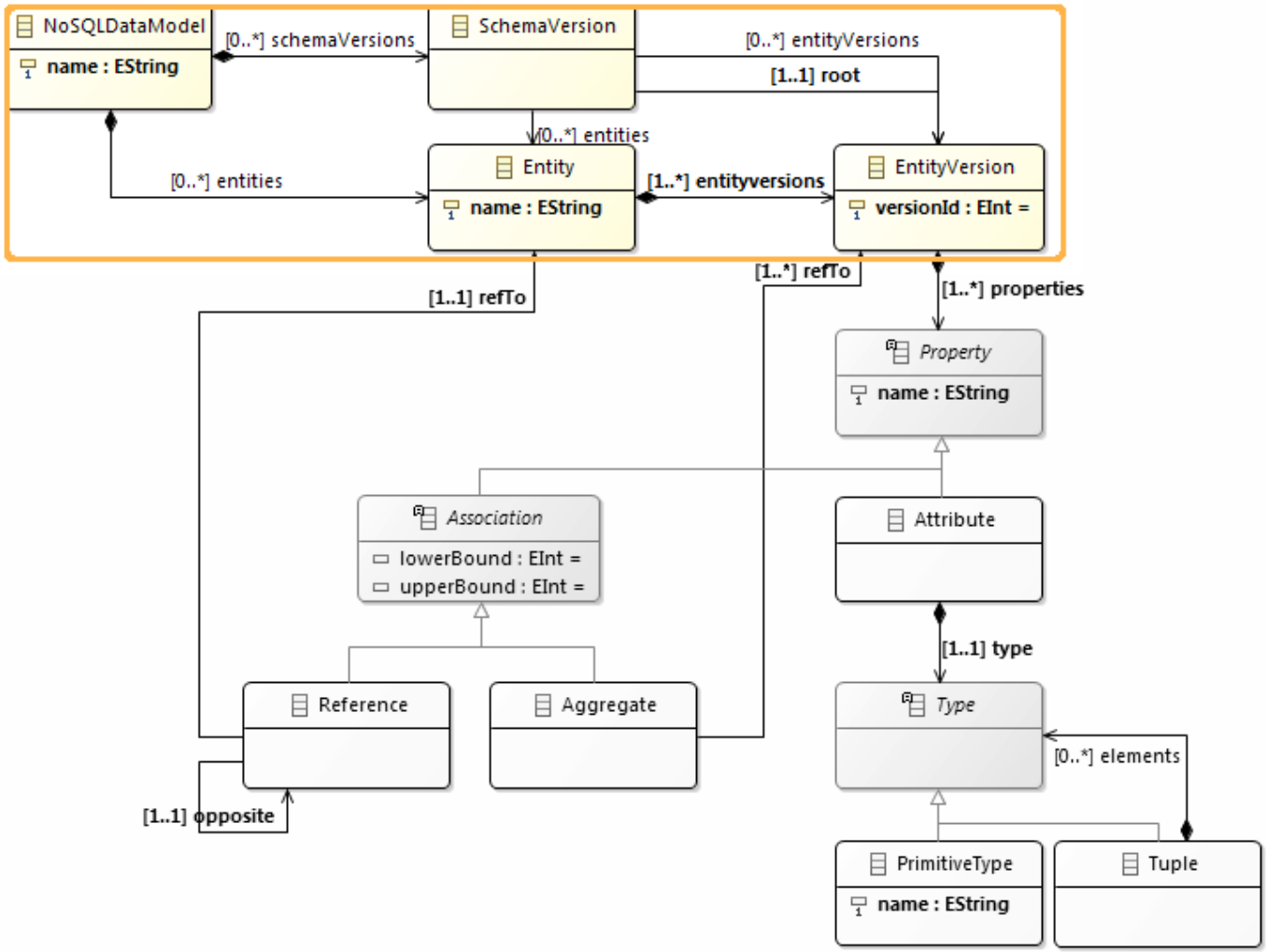
Polished metamodel (I)



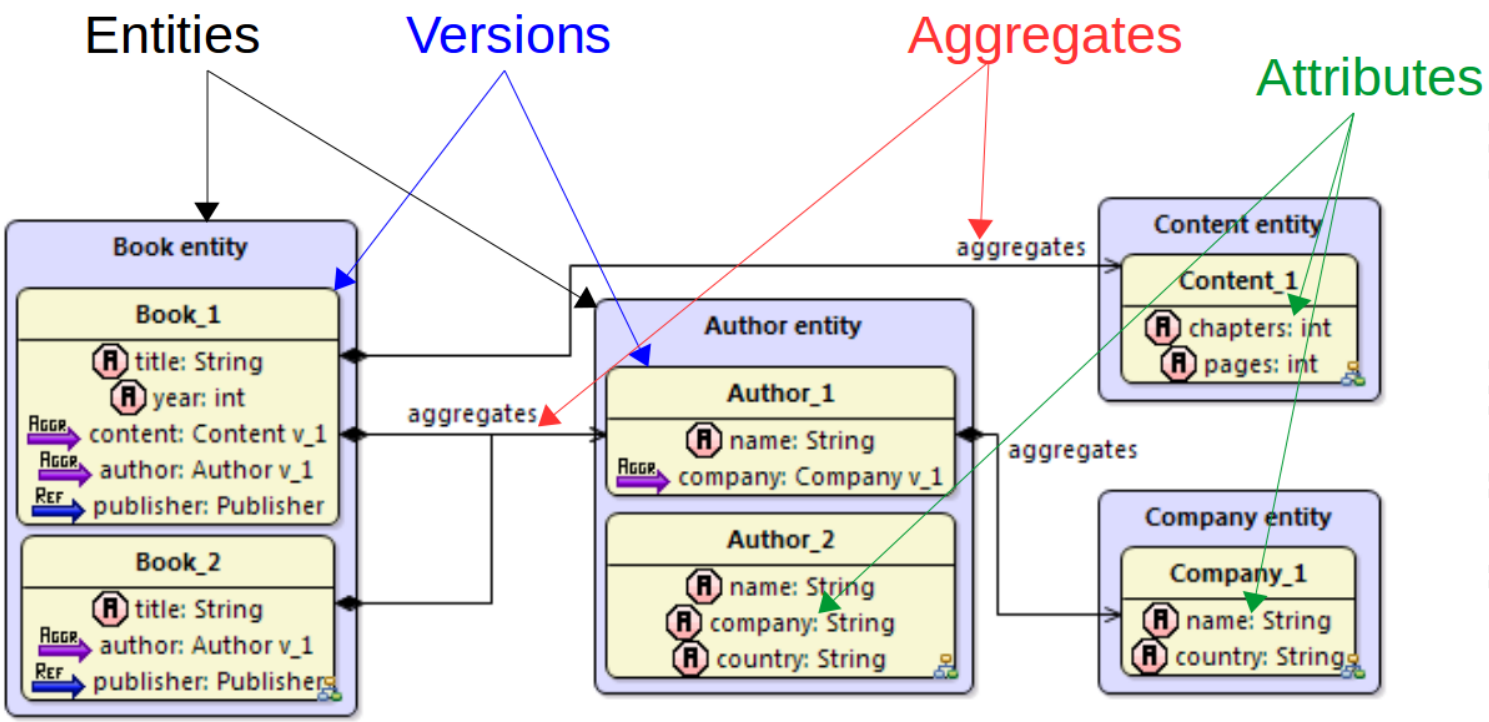
SiriusCON 16

UNIVERSIDAD DE MURCIA

Polished metamodel (II)

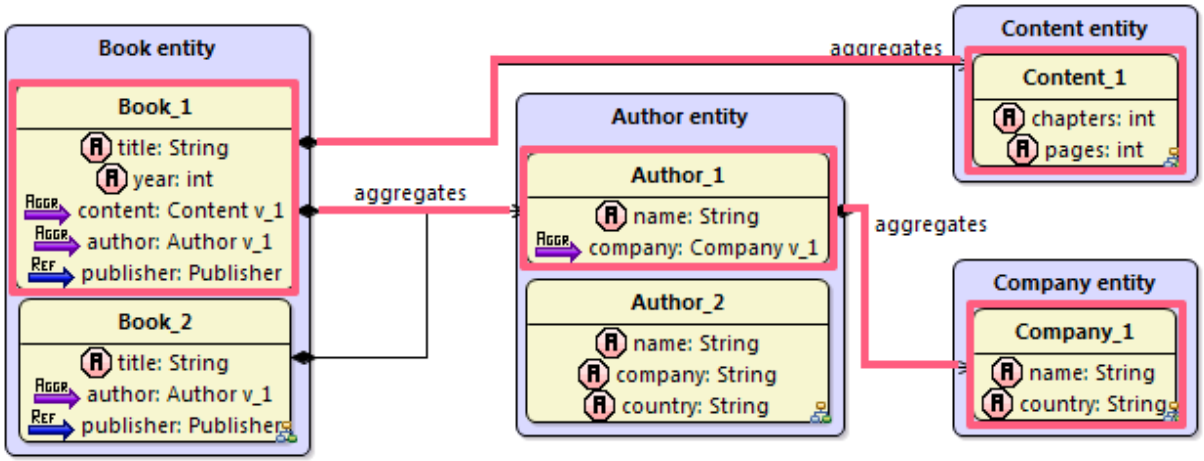


Metamodel elements



Schema versions (I)

Schema version for Book_1



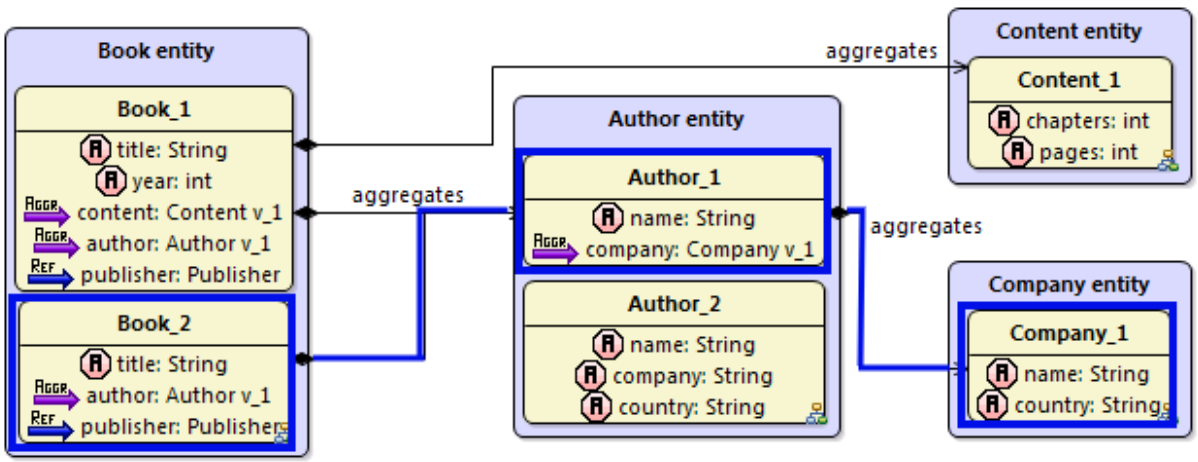

```
...
{
  "_id": "53",
  "type": "Book",
  "title": "Book_1",
  "hasContent":
  [{
    "_id": "64",
    "type": "Content",
    "chapters": 3,
    "pages": 17
  }],
  "hasAuthors":
  [{
    "_id": "155",
    "type": "Author",
    "name": "Author_155",
    "hasCompany":
    [{
      "_id": "123",
      "type": "Company",
      "name": "Company_123",
      "country": "Country_1"
    }]
  }]
}
...
```

SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema versions (II)

Schema version for Book_2



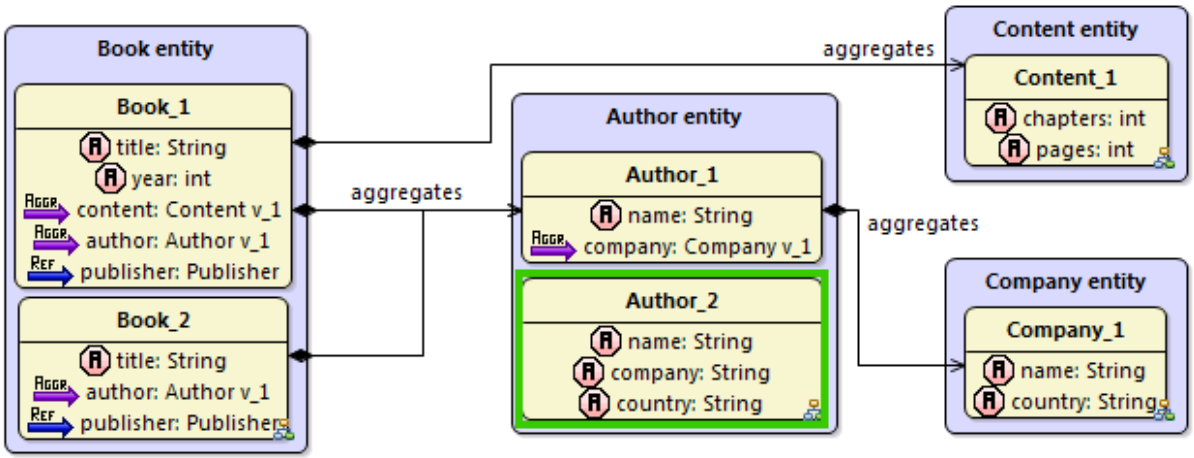

```
...
{
  "_id": "89",
  "type": "Book",
  "title": "Book_2",
  "hasAuthors":
  [{
    "_id": "77",
    "type": "Author",
    "name": "Author_77",
    "hasCompany":
    [{
      "_id": "61",
      "type": "Company",
      "name": "Company_61",
      "country": "Country_2"
    }]
  }]
}
...
```

SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema versions (III)

Schema version for Author_2




```
...
{
   "_id": "99",
   "type": "Author",
   "name": "Author_2",
   "company": "Company_3",
   "country": "Country_3"
}
...
```

Index

SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization

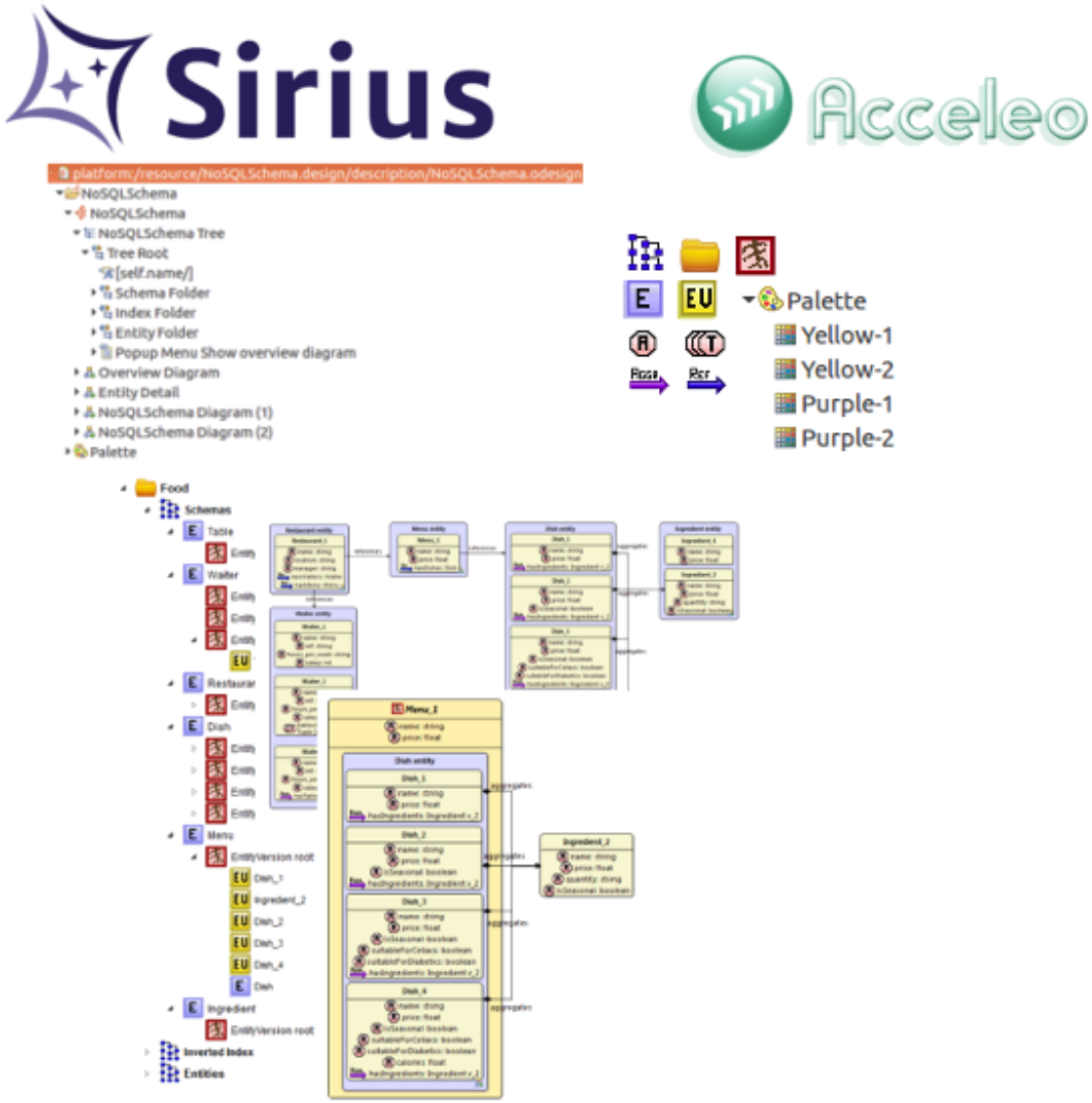
Why using Sirius?

- Views we need:
  - General tree view
  - Global schema view
  - Schema version views
  - Entity detail view
- Ability to create different viewpoints and navigability mechanisms between them
- Automatic generation of an embedded editor in the Eclipse environment
- Extensibility

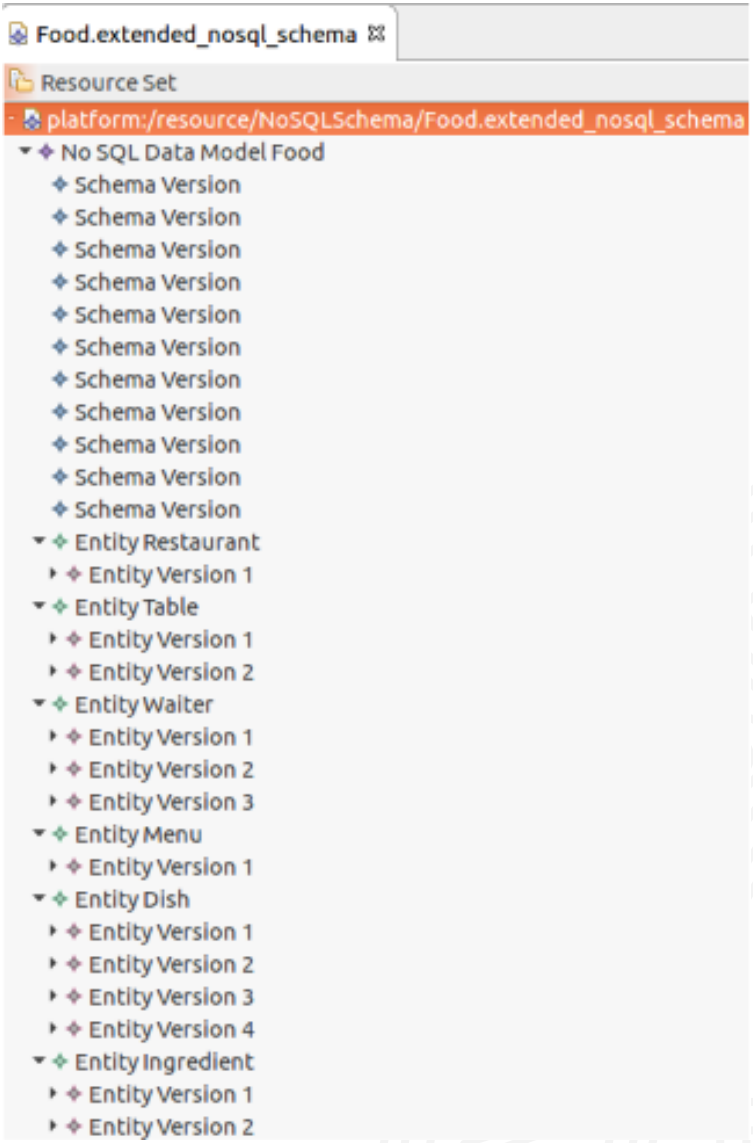
SiriusCON|16

UNIVERSIDAD DE MURCIA

Design process



Schema visualization - Input model

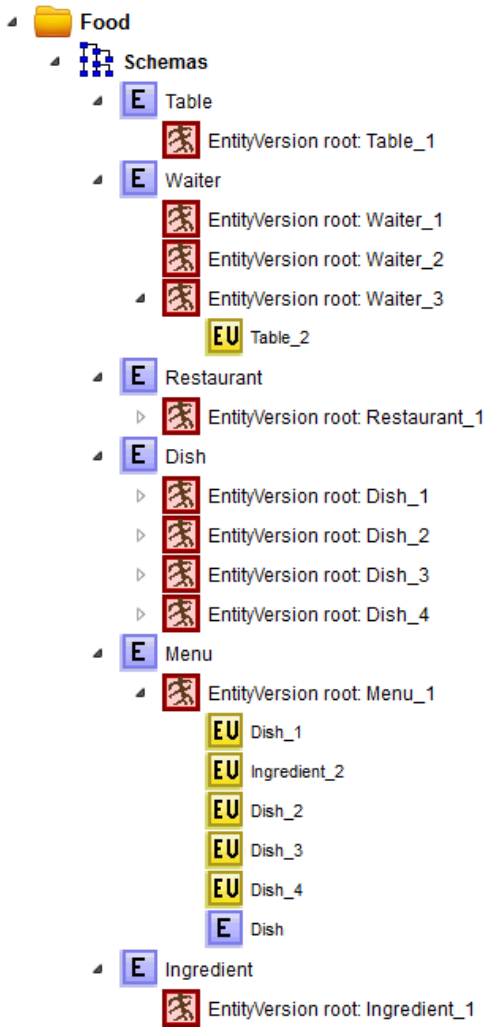


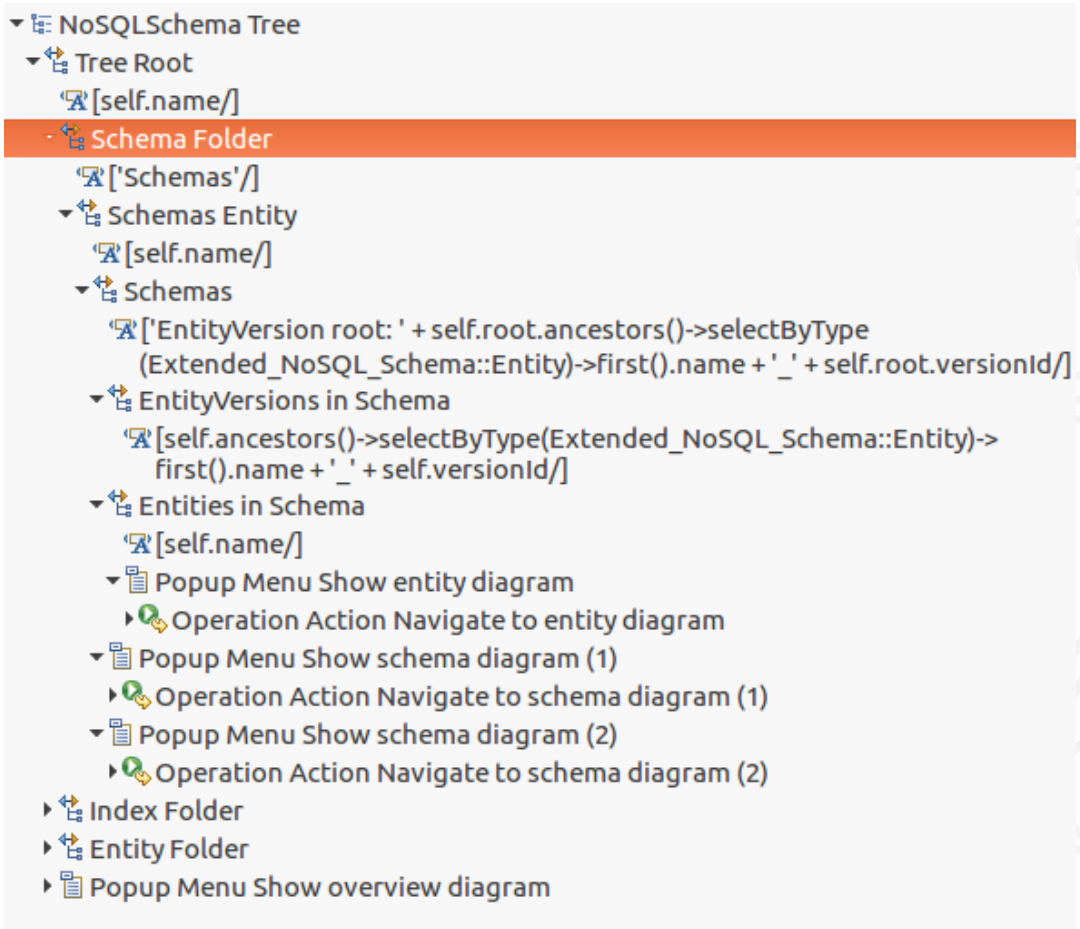
SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization - Tree viewpoint (I)

Schema versions grouped by Entity



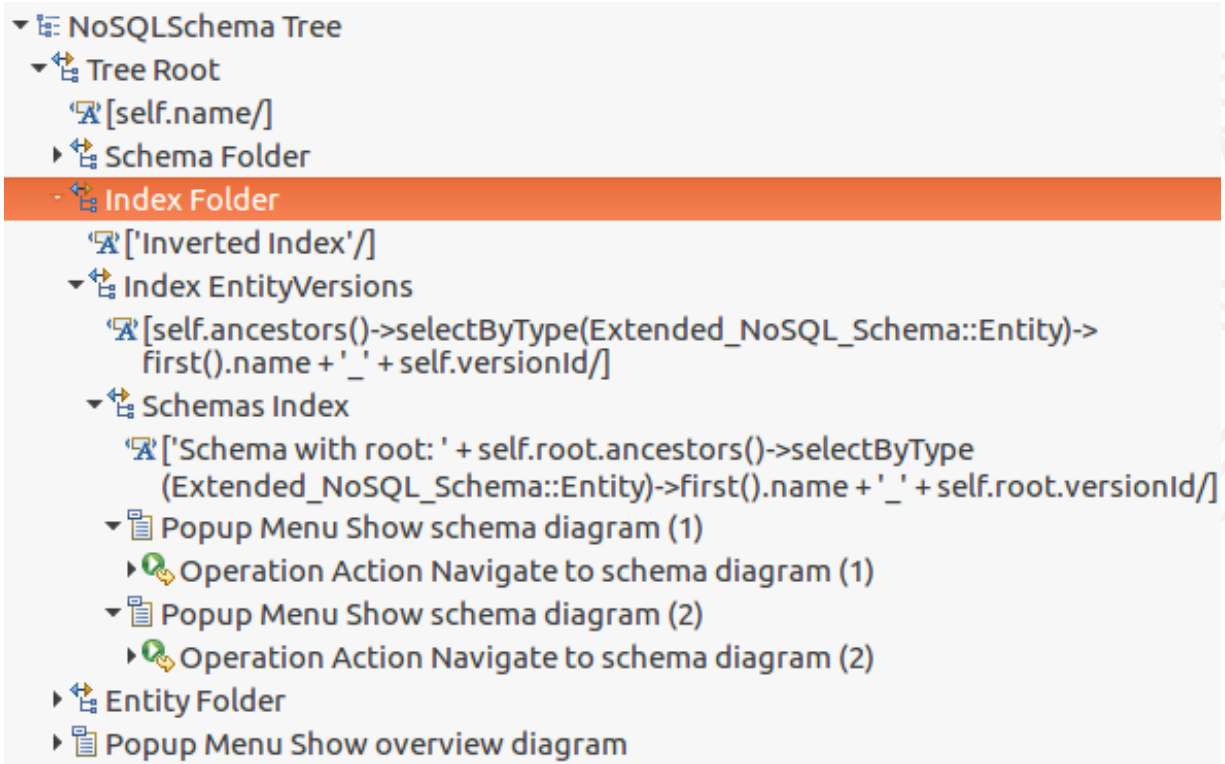


SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization - Tree viewpoint (II)

Schema versions grouped by Version

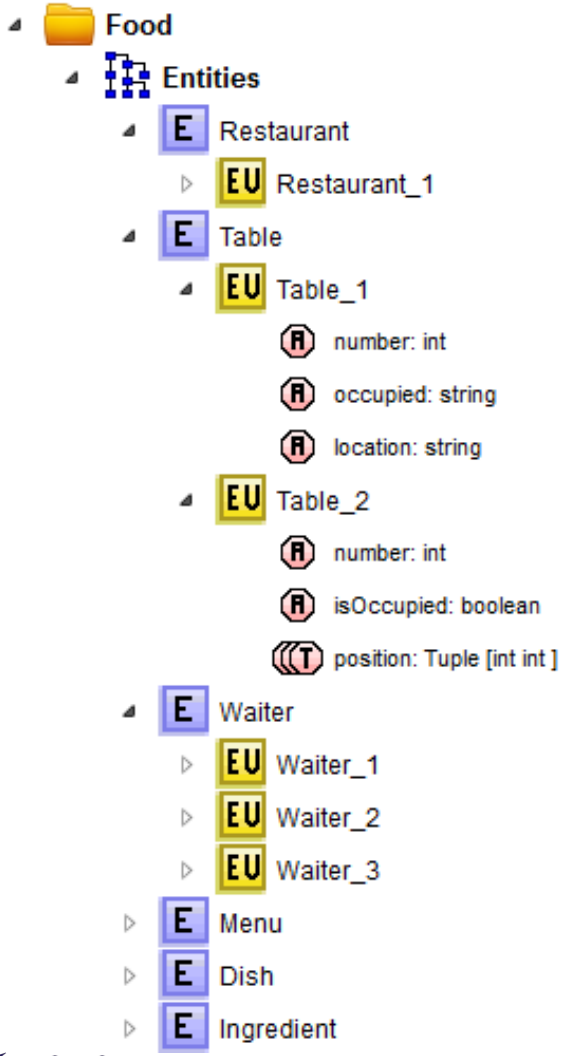


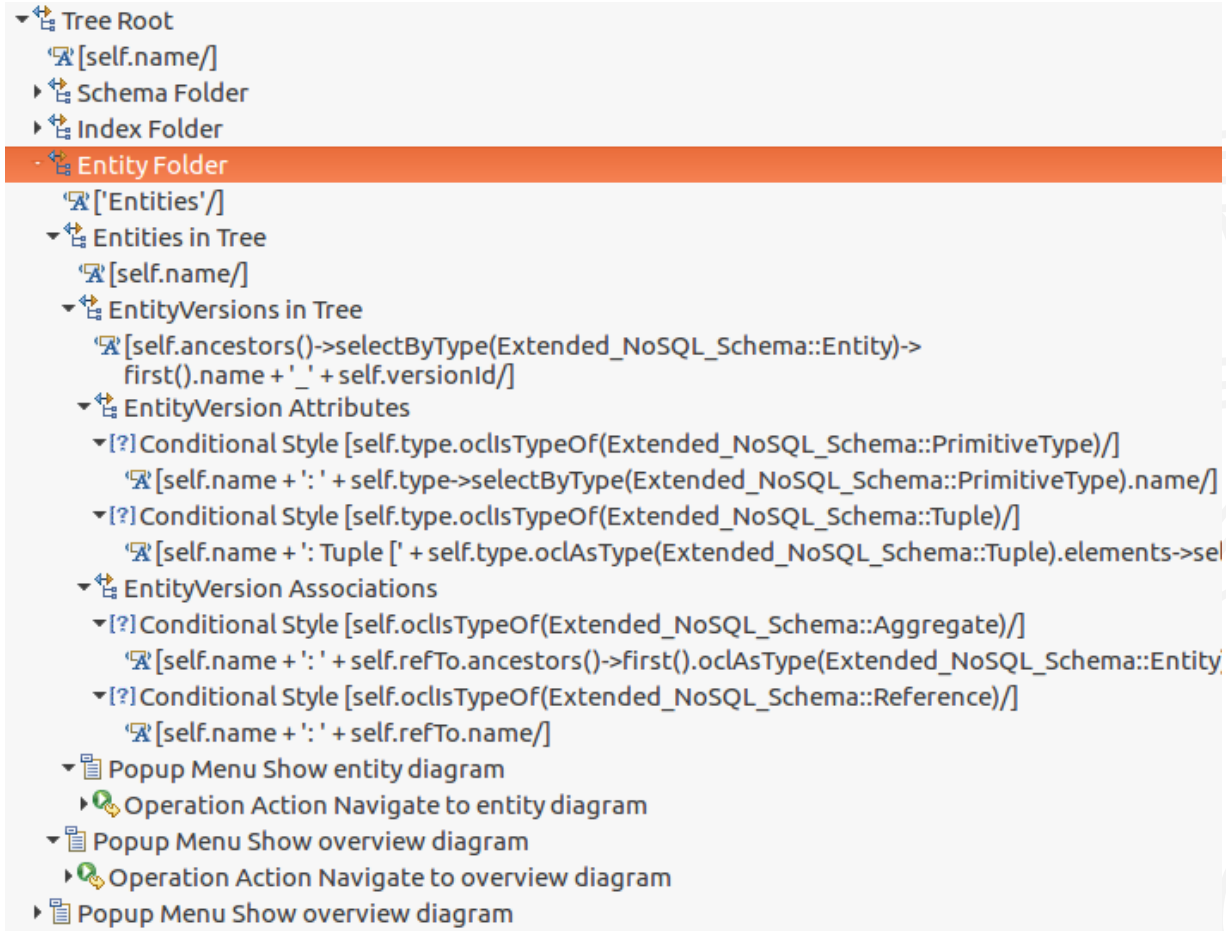
SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization - Tree viewpoint (III)

List of Entities, Versions and Properties



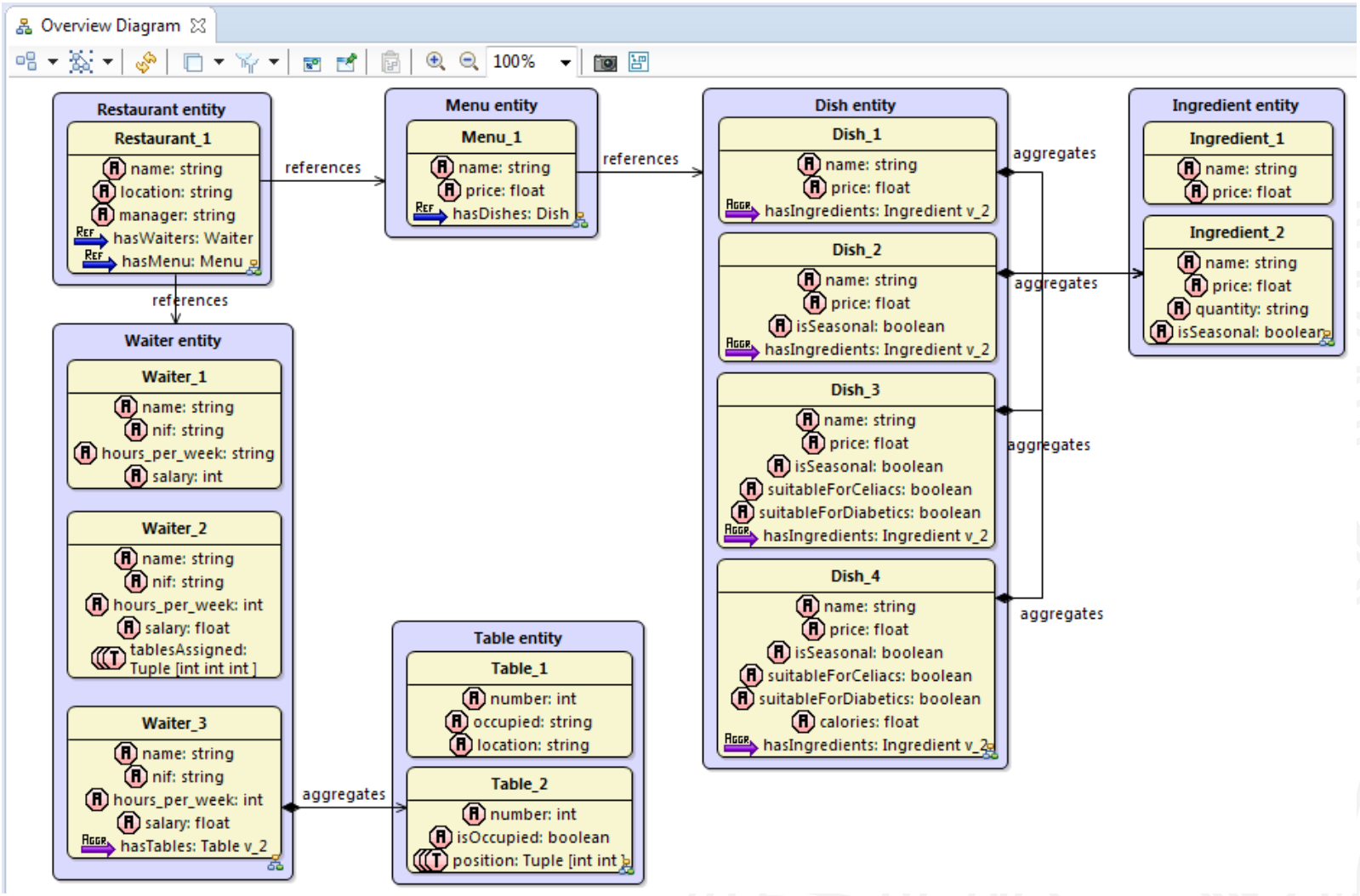


SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization - Global schema (I)

Diagram of Entities, Versions and Properties

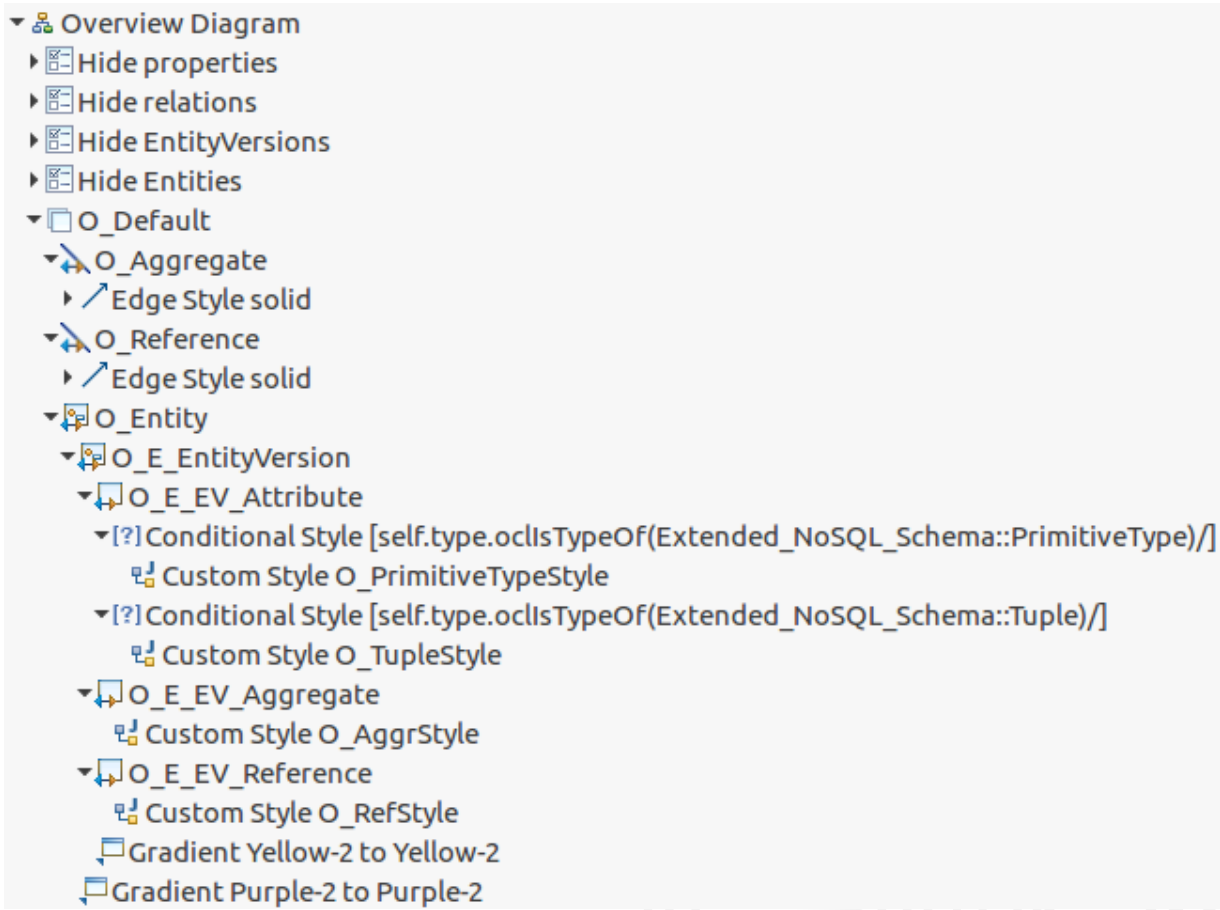


SiriusCON 16

UNIVERSIDAD DE MURCIA

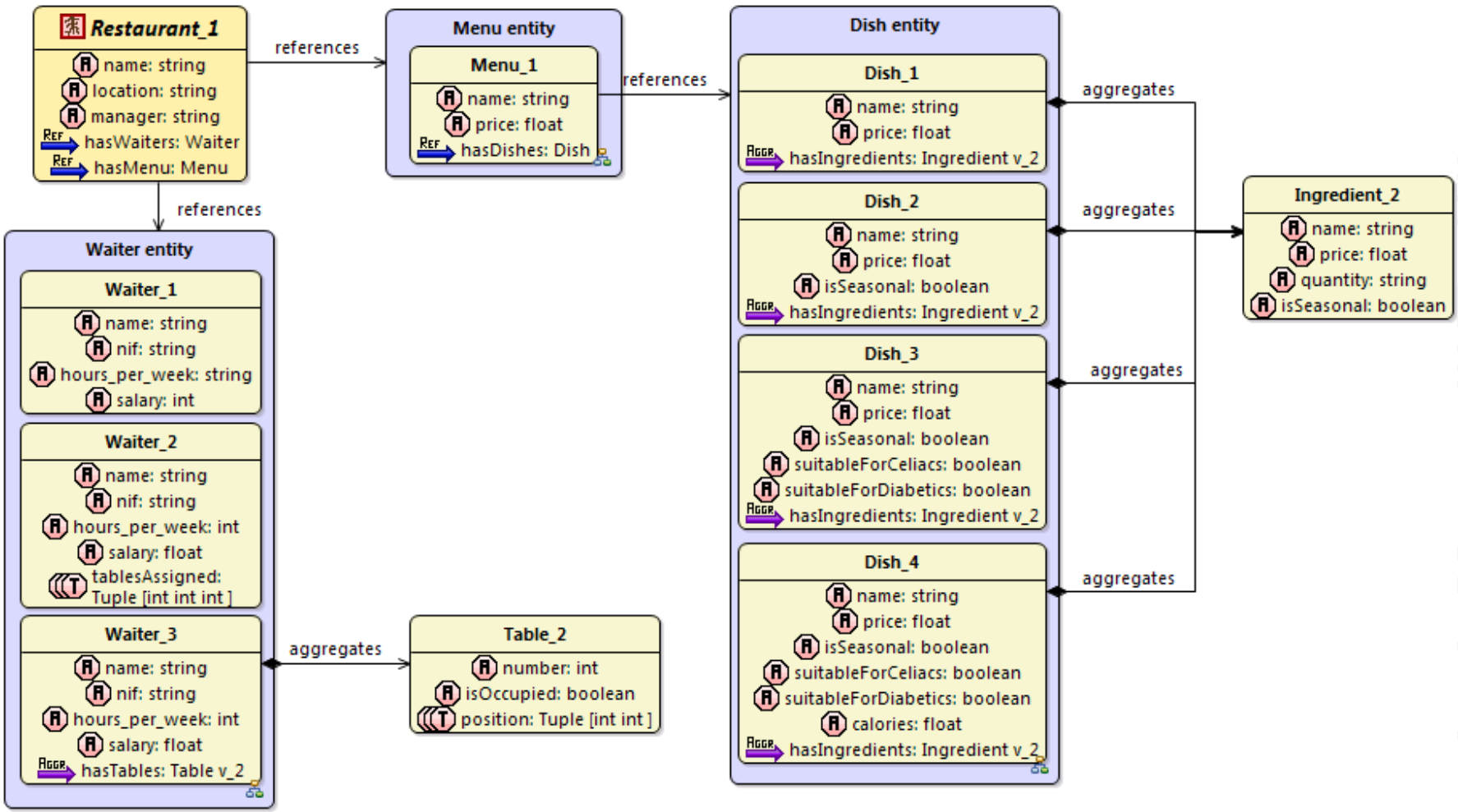
Schema visualization - Global schema (II)

Implementation of the Global schema



Schema visualization - Schema version (I)

Schema version root and its associations

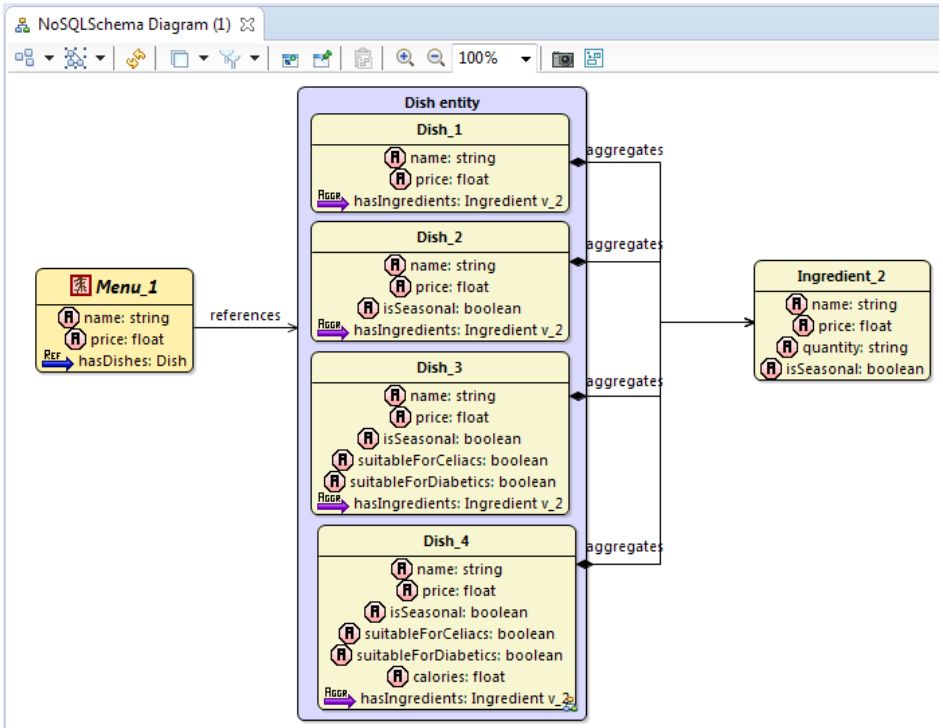


SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization - Schema version (II)

Implementation of the Schema view

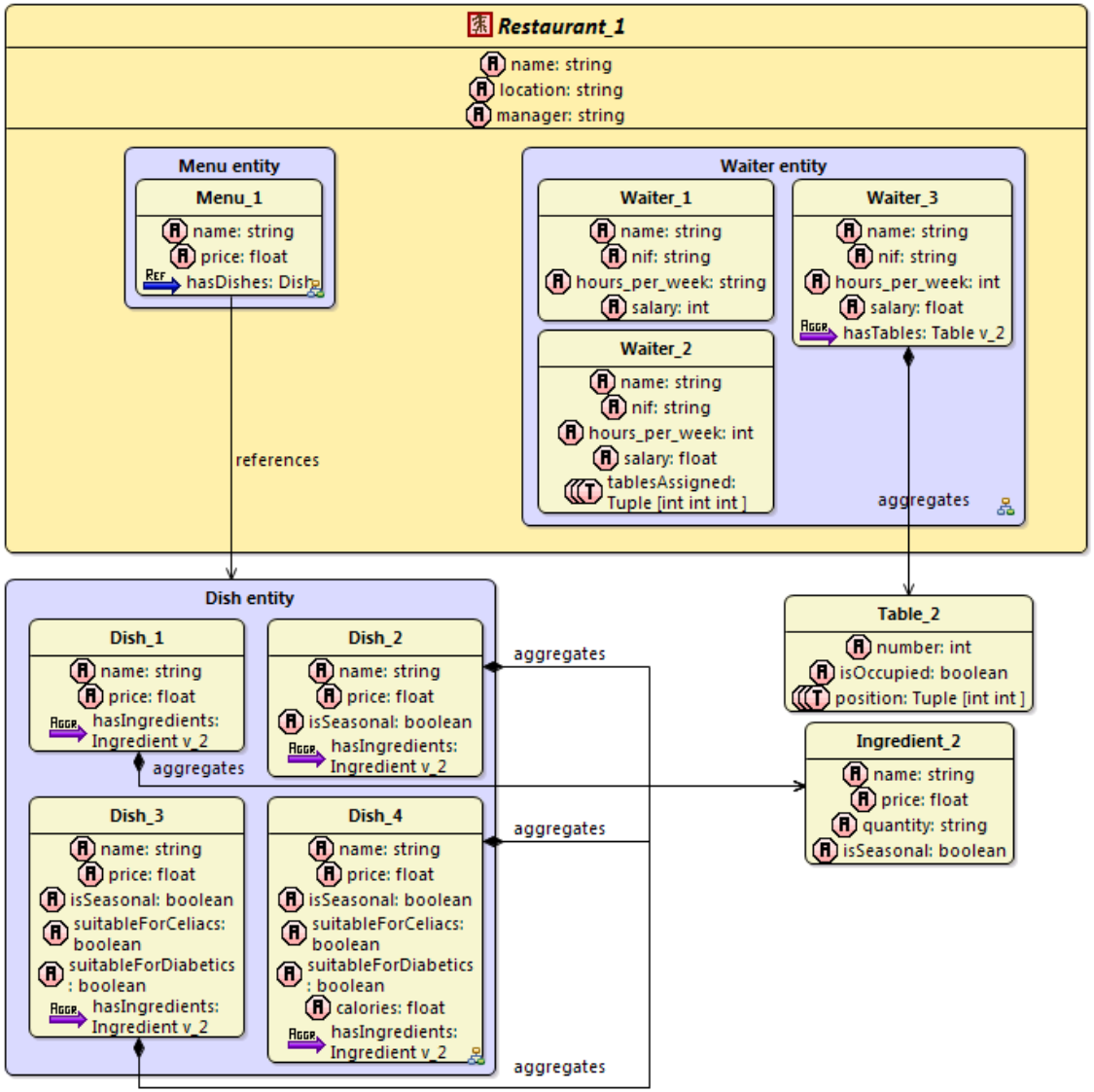


SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization - Schema version (III)

Variation with embedded direct associations

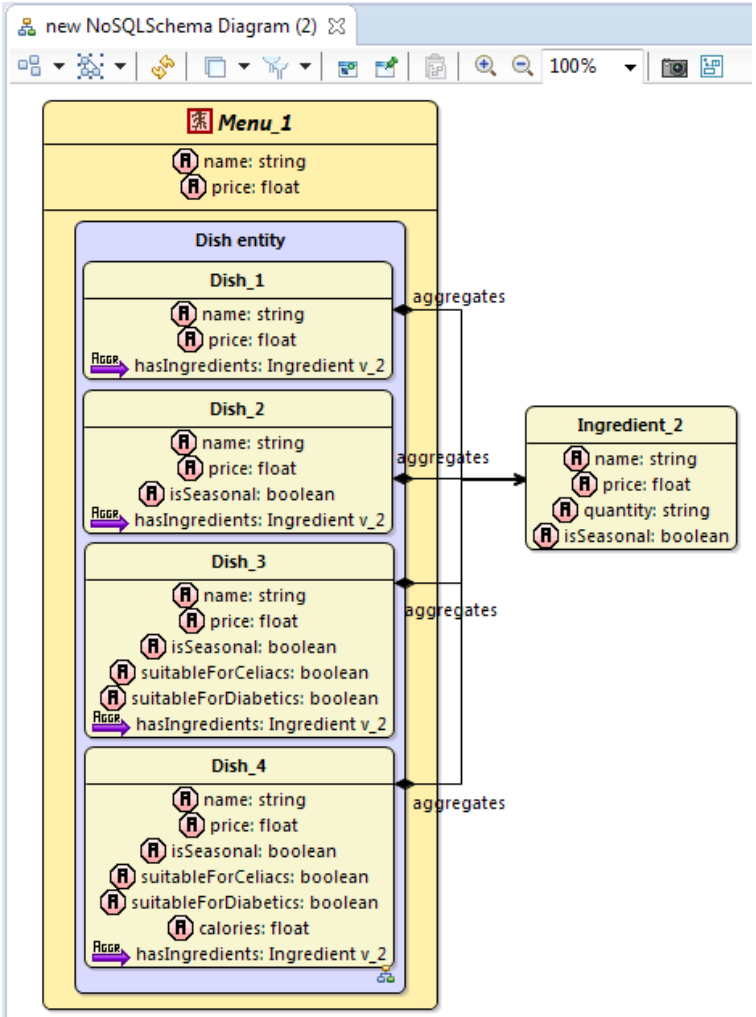


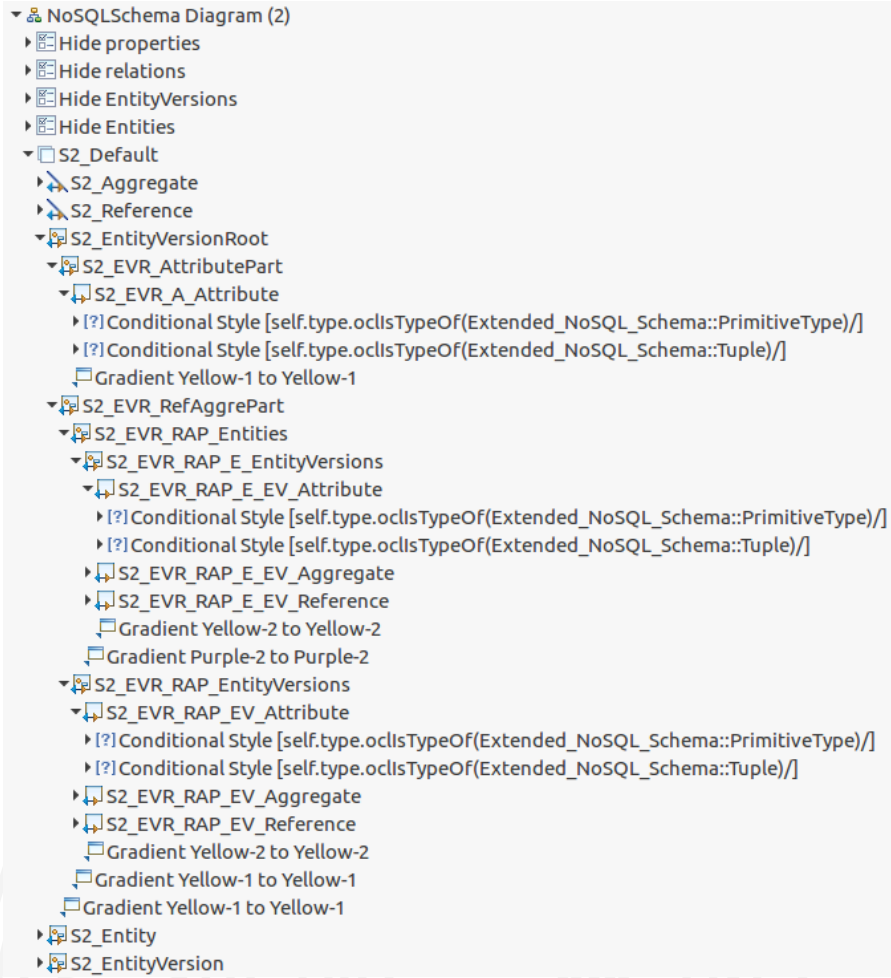
SiriusCON 16

UNIVERSIDAD DE MURCIA

Schema visualization - Schema version (IV)

Implementation of the Schema view



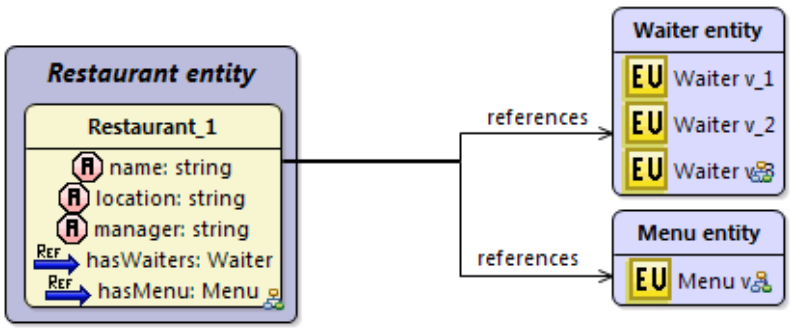


SiriusCON 16

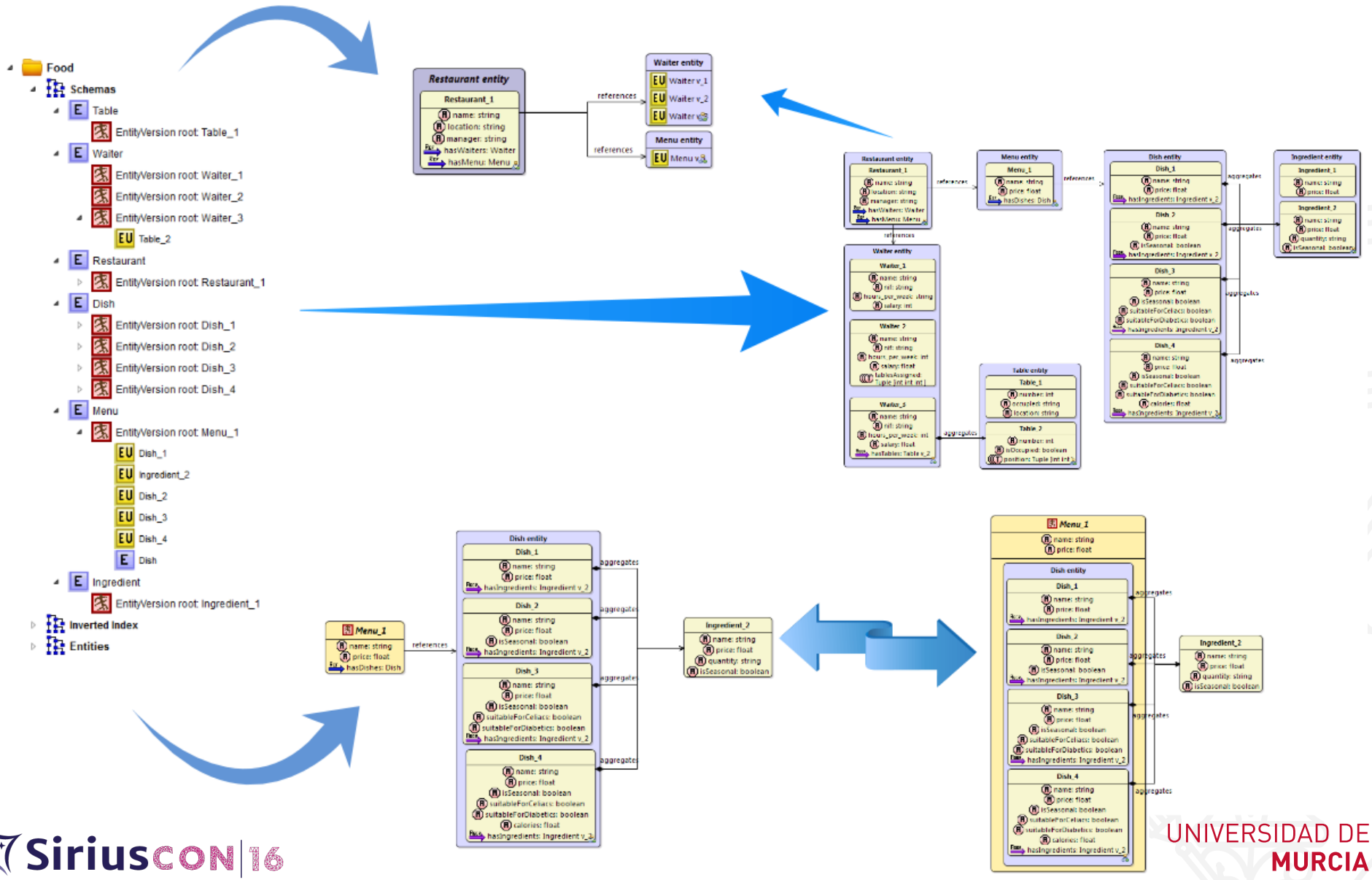
UNIVERSIDAD DE MURCIA

Schema visualization - Entity detail

Entity detail with its Versions associations



Navigation between views



Index

SiriusCON|16

UNIVERSIDAD DE MURCIA

To sum up (I)

About Sirius...

- ✔ Built in the Eclipse Modeling Framework
- ✔ Easily deployable along with the metamodel into plugins
- ✔ Different viewpoints, layers and customization options
- ✔ Less development time and easier to extend
- ✘ AQL might be quite tricky sometimes when defining actions
- ✘ Learning curve for creating complex examples may be too harsh
- ? Language/model maintenance and evolution
- ? Managing large input models

SiriusCON 16

UNIVERSIDAD DE MURCIA

To sum up (II)

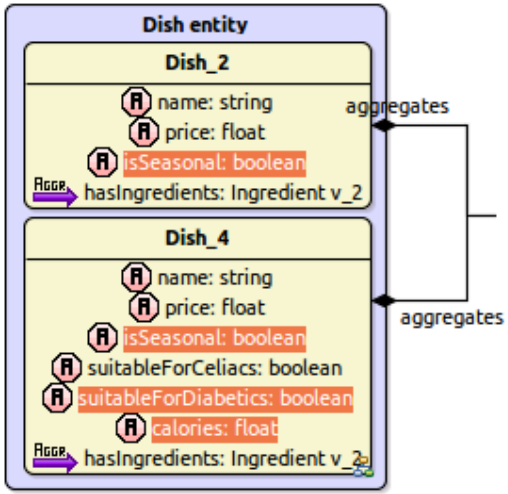
About the tool...

- We developed one of the first approaches for NoSQL visualization considering versions
- Tool to visualize NoSQL database schemas and schema versions with Sirius
  - Tree viewpoint and global schema viewpoint
  - Schema version and entity detail viewpoint
- ...but there is still a lot of work to do:
  - Polish the NoSQL_Schema metamodel
  - Improve the Sirius viewpoint definitions
  - Implement the editor functionality

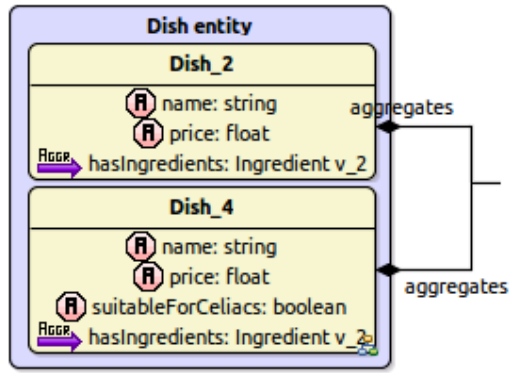
SiriusCON|16

UNIVERSIDAD DE MURCIA

Future work (I) - DB alterations



Removing attributes




```
{
    "_id": 45,
    "type": "Dish",
    "name": "value_190",
    "price": 0.43,
    "isSeasonal": true,
    "hasIngredients": ...
},
{
    "_id": 50,
    "type": "Dish",
    "name": "value_77",
    "price": 53.84,
    "isSeasonal": false,
    "suitableForCeliacs": false,
    "suitableForDiabetics": true,
    "calories": 7.16,
    "hasIngredients": ...
},
```


```
hbase(main):001:0> delete 'food_db', '45', ...
                       ... 'isSeasonal'
hbase(main):002:0* delete 'food_db', '50', ...
                       ... 'isSeasonal', 'suitableForDiabetics', 'calories'
```

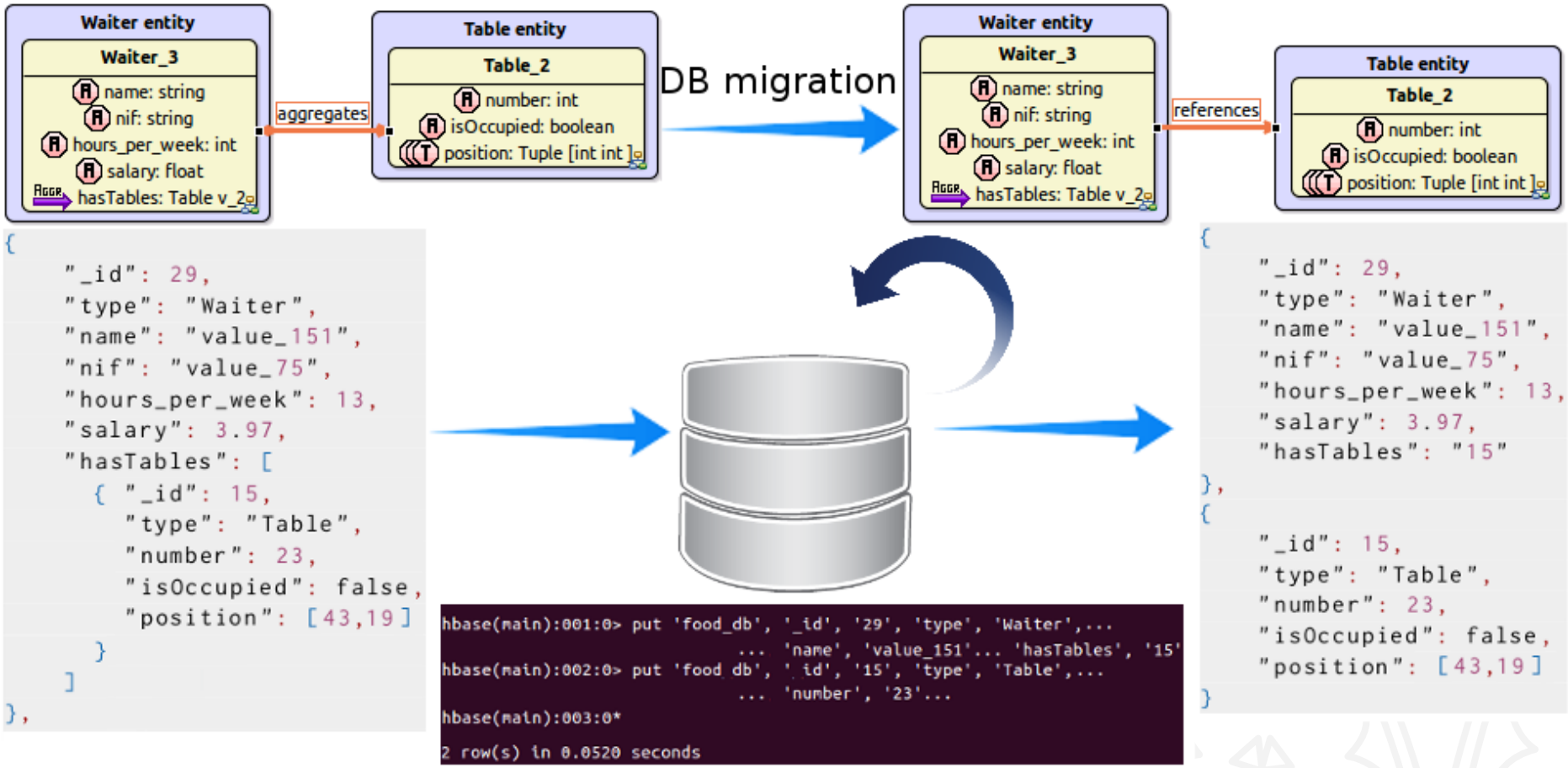
Generate transformation scripts for the DB


```
{
    "_id": 45,
    "type": "Dish",
    "name": "value_190",
    "price": 0.43,
    "hasIngredients": ...
},
{
    "_id": 50,
    "type": "Dish",
    "name": "value_77",
    "price": 53.84,
    "suitableForCeliacs": false,
    "hasIngredients": ...
},
```


SiriusCON 16

UNIVERSIDAD DE MURCIA

Future work (II) - DB migration



Generate migration scripts for the DB

SiriusCON 16

UNIVERSIDAD DE MURCIA

alberto.hernandez1@um.es https://github.com/Soltari https://linkedin.com/in/albertohc

SiriusCON 16

UNIVERSIDAD DE MURCIA