

# COHESIVE Information system



Adriano Di Pasquale

COHESIVE WP4 - [a.dipasquale@izs.it](mailto:a.dipasquale@izs.it)



## Task 4.1. Objectives

The OHEJP COHESIVE project aims to improve efficiency on outbreak management, surveillance and risk assessments in a One-Health approach at Member State level.

The goal can be achieved by

**integrating information on pathogens  
from the human and veterinarian side,  
present at  
Member State level**



## Task 4.1. Information

Information under consideration in COHESIVE are:

- **WGS data** produced by analysis of pathogens by the Member State laboratories
- **Metadata**, such as the minimum epi-data associated with a pathogen.

Moreover, additional useful metadata sources will be considered



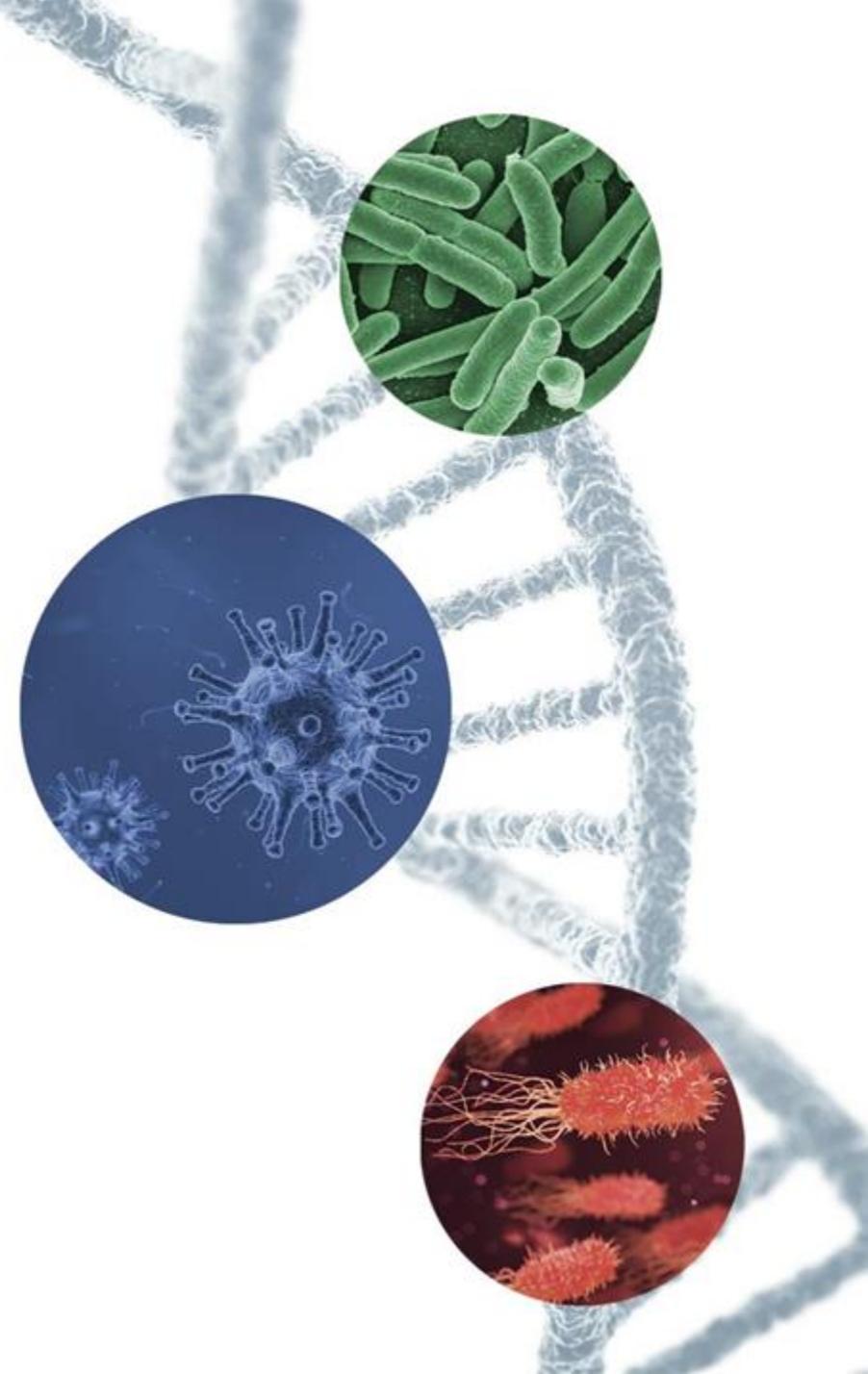
## Task 4.1. Information: WGS data

### ■ Analysis results

- Raw data
- De novo assembly
- Typing (e.g., MLST, cgMLST)
- Etc.

### ■ Metadata of analysis

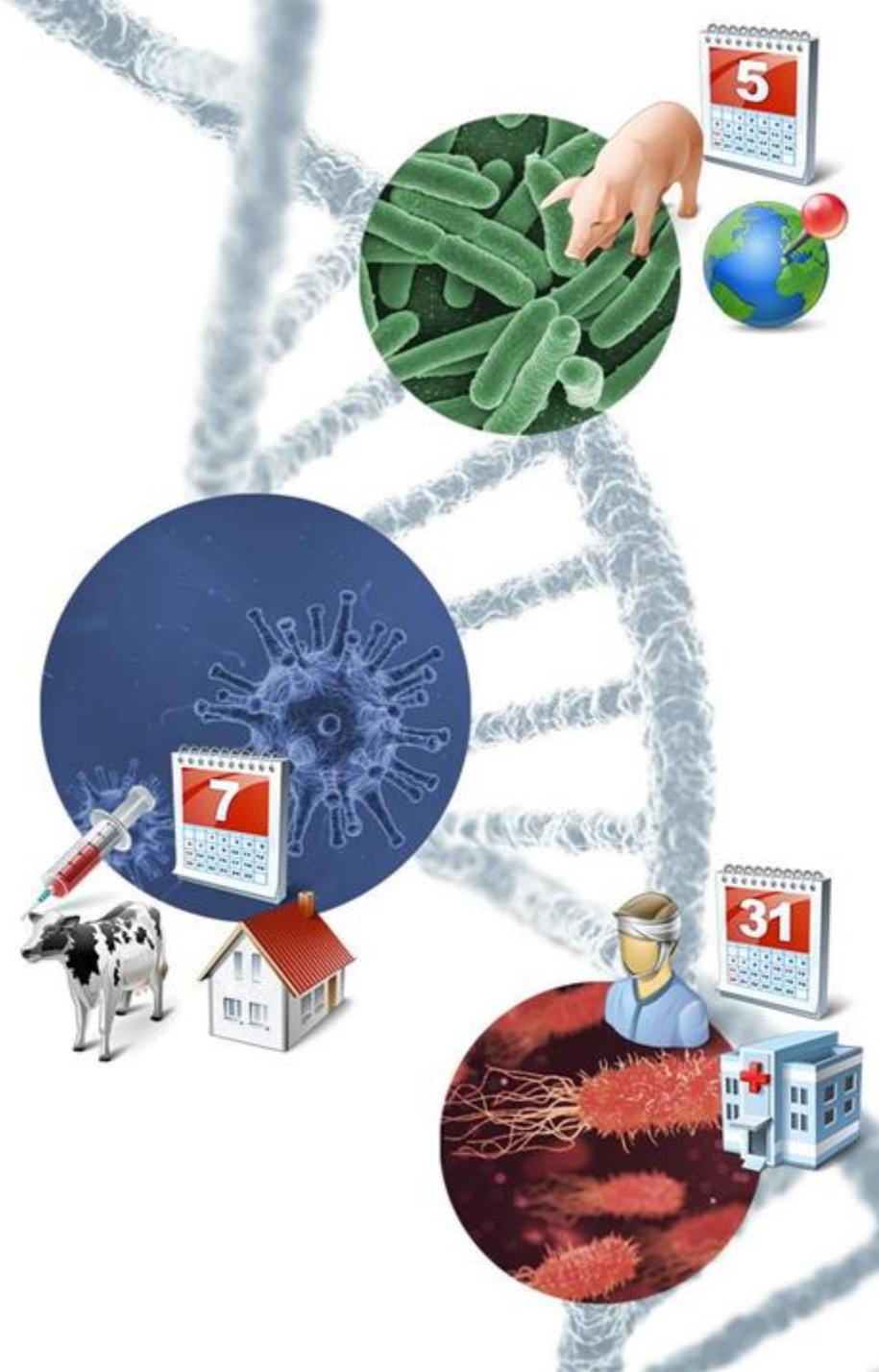
- Format
- Tools: parameters and versions
- Quality





## Task 4.1. Information: Metadata:

- **What:** pathogen species
- **Matrix** (source of isolation):
  - cheese, blood, water
- **When, Where:**
  - sample collection date and place





# Task 4.1. Feasibility study

.



## Task 4.1. Feasibility study

**The main activity is to analyze at MS level the feasibility of such integration.**

Countries under evaluation	Species for pilot
Italy	Listeria
The Netherlands	Listeria and HEV
Norway	Listeria

On each Member State,  
the integration of national systems will be analyzed.



## Task 4.1. Steps of the feasibility study at MS level

- Choice of pathogen species
- List of the **involved organizations**
- List of all the useful **information sources**,
- A **prototype system** is used to let the Member State evaluate the expected improvements:
  - a) Evaluate **integration** of the systems
  - b) Evaluate the possibility to provide **EU harmonized output**.



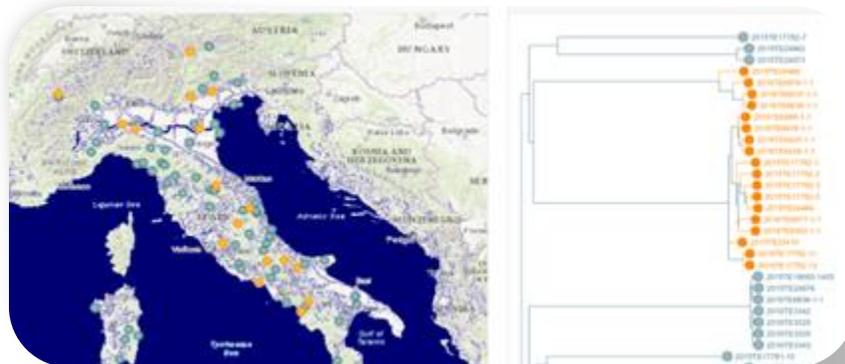
## Task 4.1. Steps of the feasibility study at MS level

- Choice of pathogen species
- List of the **involved organizations**
- List of all the useful **information sources**,
- **A prototype system** is used to let the Member State evaluate the expected improvements:
  - a) Evaluate **integration** of the systems
  - b) Evaluate the possibility to provide **EU harmonized output**.



# Task 4.1. Prototype Information System

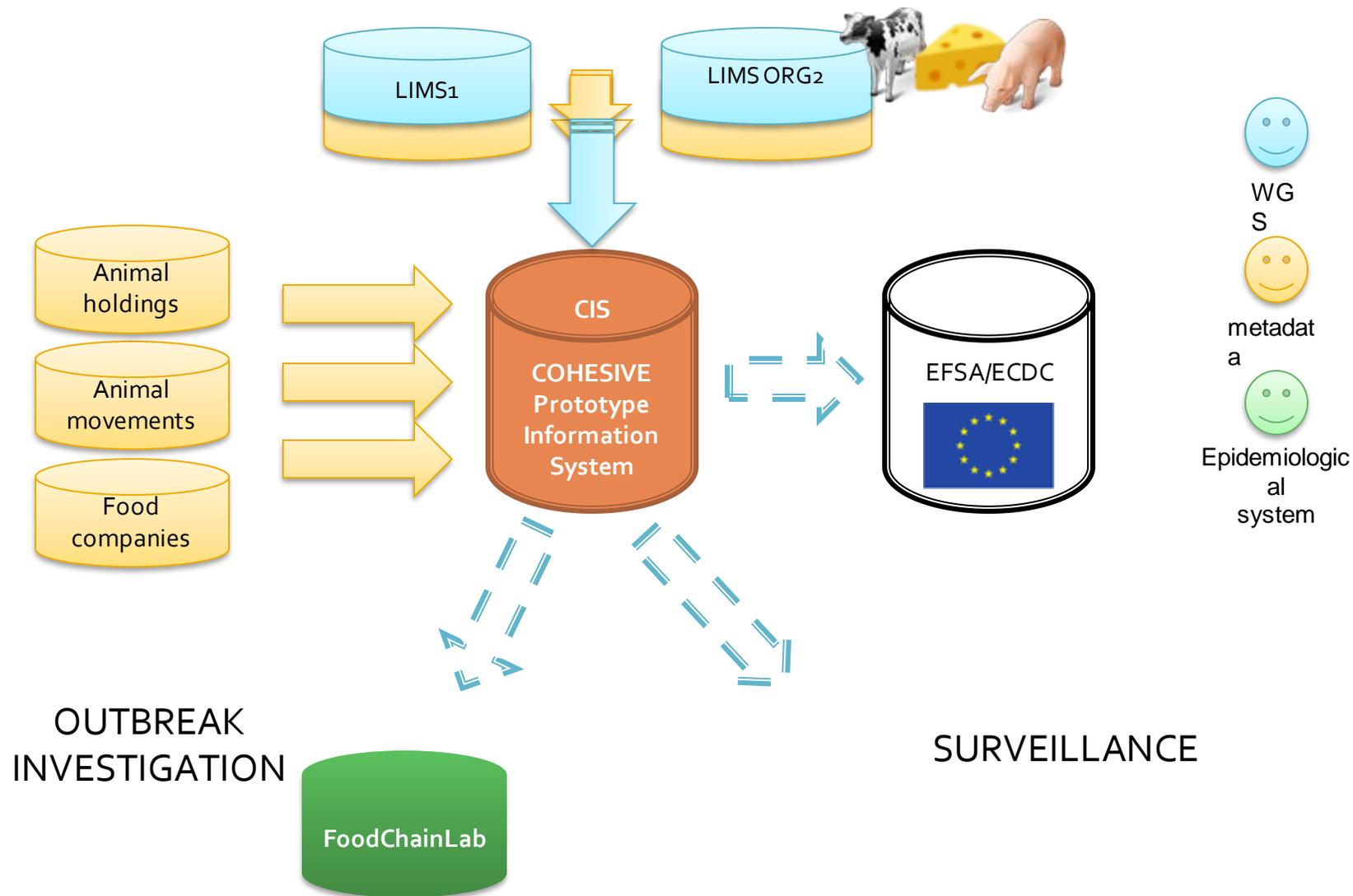
- A WEB based information system with **HIGHLY FLEXIBLE** database schema
- Provide a concrete example of the systems integration
- Just for evaluation of improvements on outbreak investigation and surveillance.
- It is up to the Member State to implement a final system (or to decide that it is not needed).







## 4.1.2. COHESIVE prototype information system (CIS). System architecture





# CIS – data component



Based on **CMDBuild**

<http://www.cmdbuild.org/>

an open source environment for configuring highly customized applications

No software development needed for

- New tables
- New table fields
- New tables relationships
- New views
- New form fields

Any specific organization need can be easily incorporated





# CIS additional components

Data component



The **CMDBuild** system has been extended in order to:



# CIS additional components

Data component



The **CMDBuild** system has been extended in order to:

Calculation engine component



**nextflow**

Integrate a calculation node based on **Nextflow** in order to perform bioinformatics analysis



# CIS additional components

Data component



The **CMDBuild** system has been extended in order to:

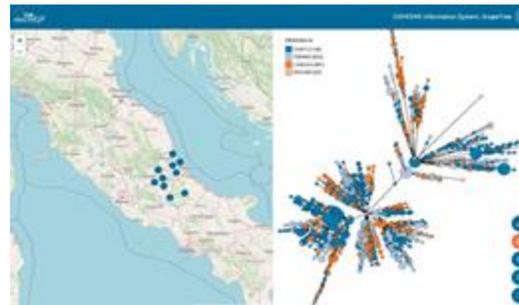
Calculation engine component



**nextflow**

Integrate a calculation node based on **Nextflow** in order to perform bioinformatics analysis

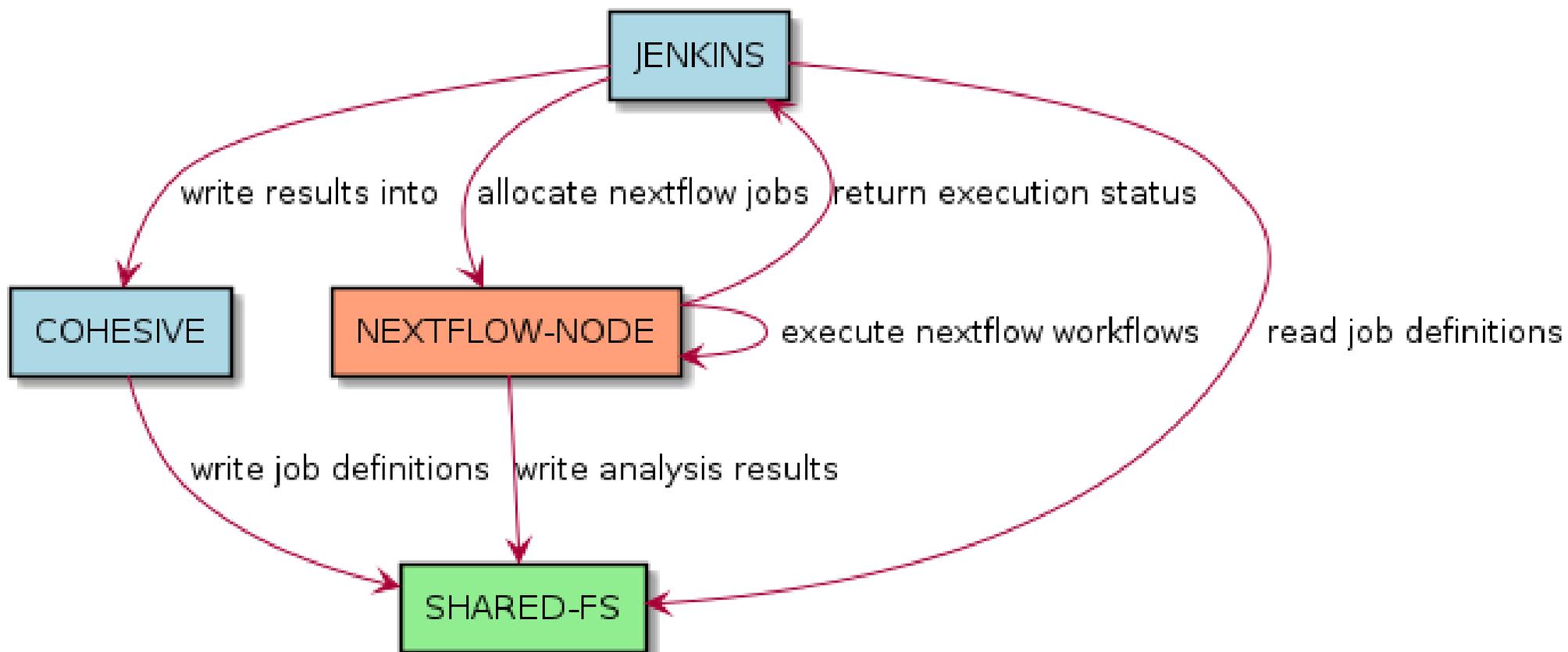
Dashboard component



Integrate dashboard combining WGS analysis with metadata and GIS (GrapeTree, Openstreetmap, Phylocanvas)



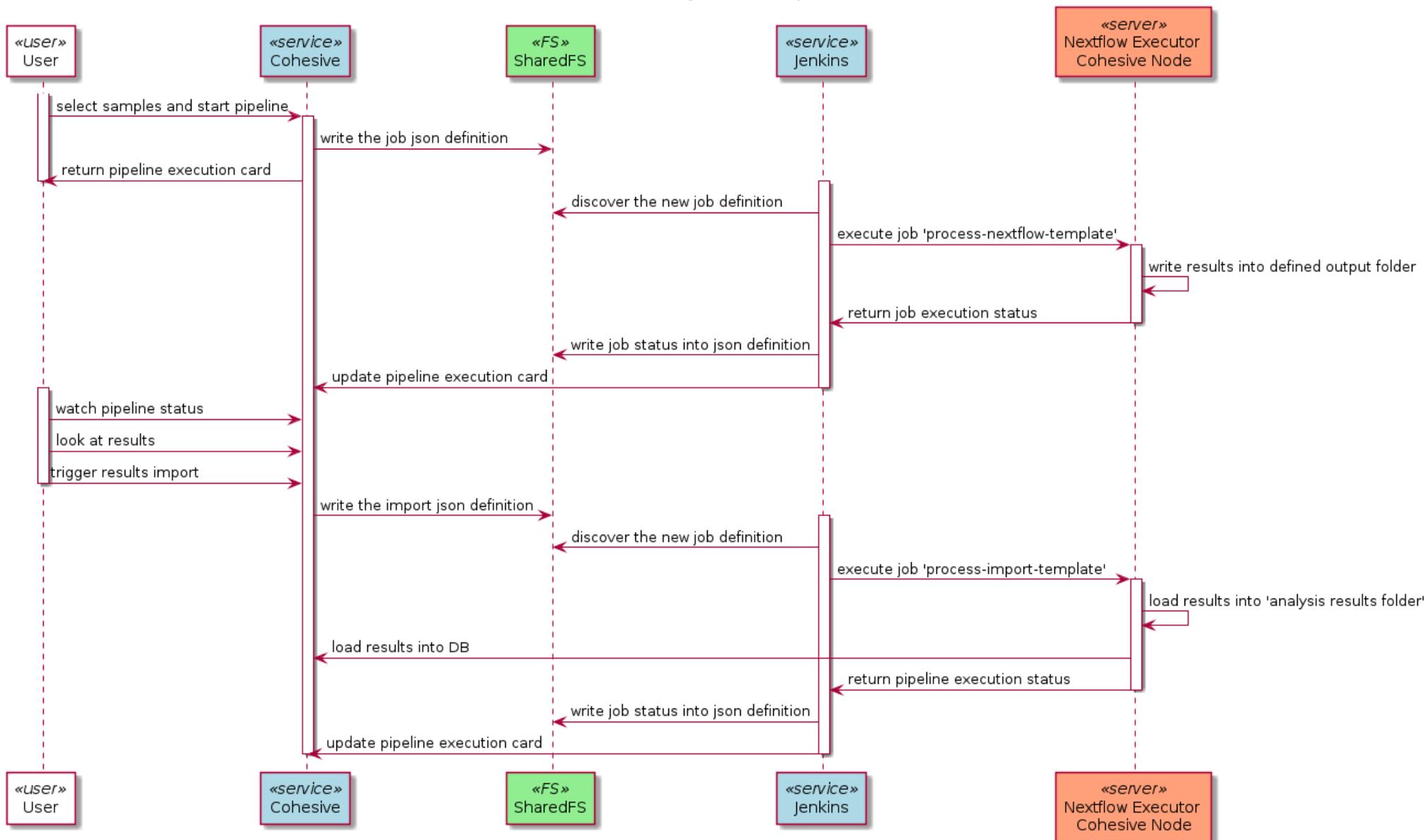
# CIS Architecture for Analysis





# CIS Architecture for Analysis

## Execute an analysis and import results

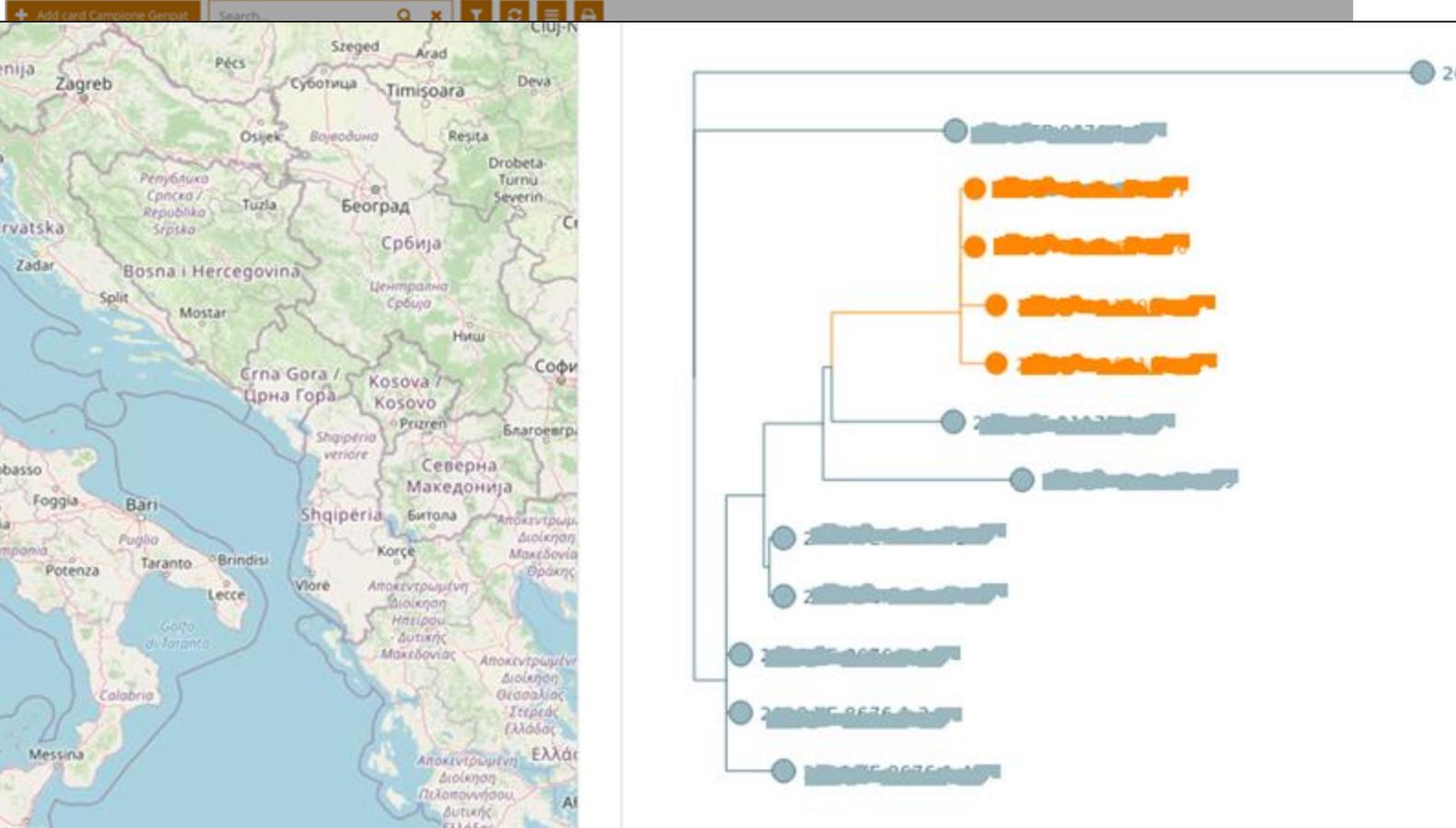








### Cards Campione Genpat



**Risultato Grape Tree**  
[Albero MST con GrapeTree - Minimum Spanning Tree in formato newick](#)  
[Albero NJ con GenPAT GIS - Neighbor joining in formato newick](#)  
[Chiamata Allelica](#)  
[Metadati](#)  
[Missing Loci](#)  
[Log ceMLST](#)

Sevo | Sequence Type | Richieder

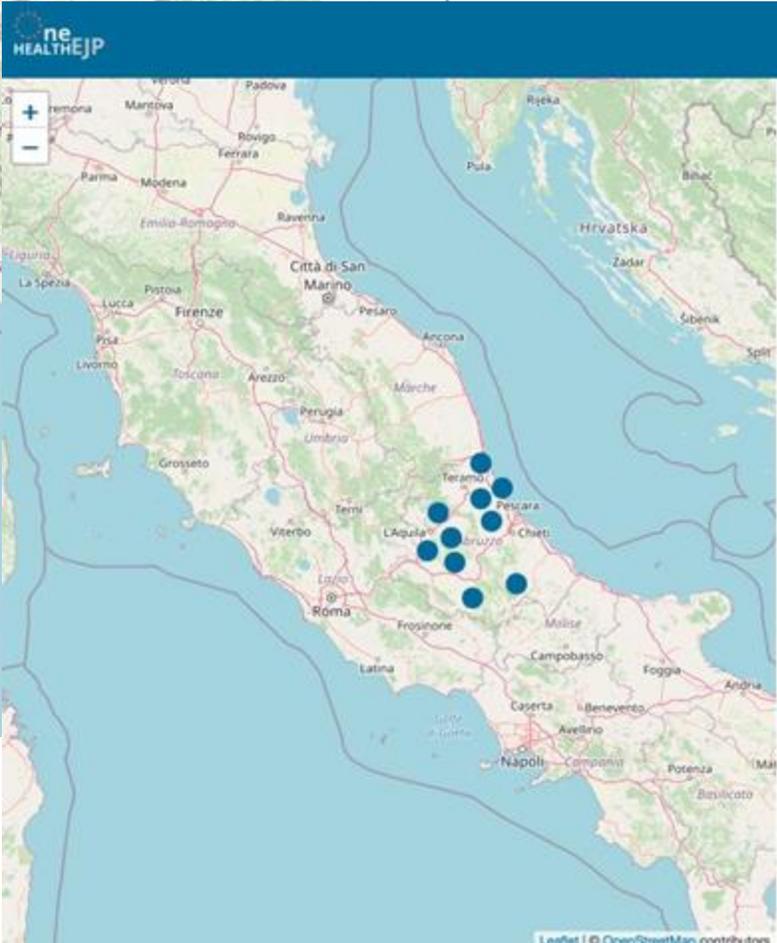
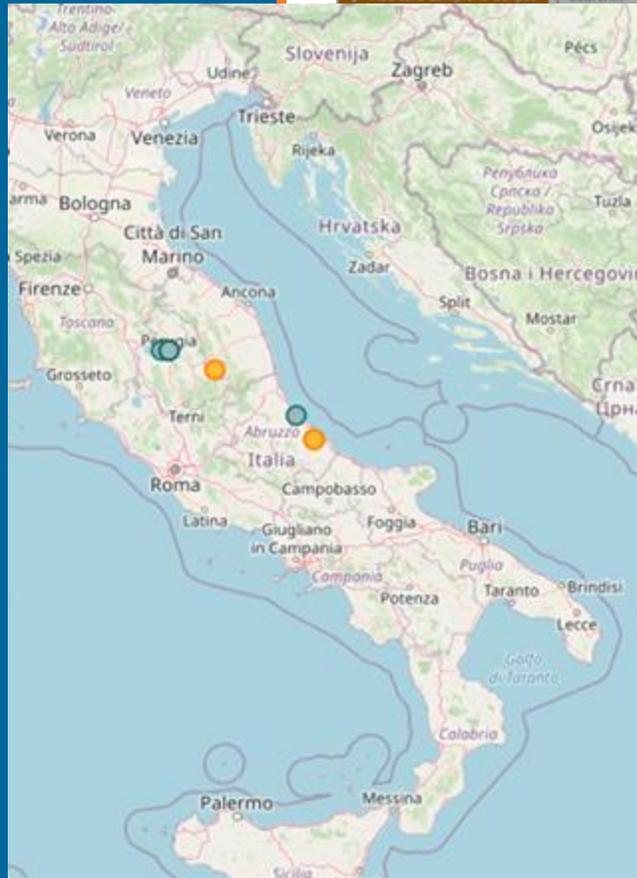
Browse...

[Dati dei Reference per il template di](#)



Cards Campione Genpat

+ Add card Campione Genpat Search [magnifying glass icon] [x icon] [T icon] [refresh icon] [list icon] [share icon]



Richieder

[x icon]

[refresh icon]

[search icon]

[refresh icon]



# Analysis: New SARS-CoV-2 clustering method

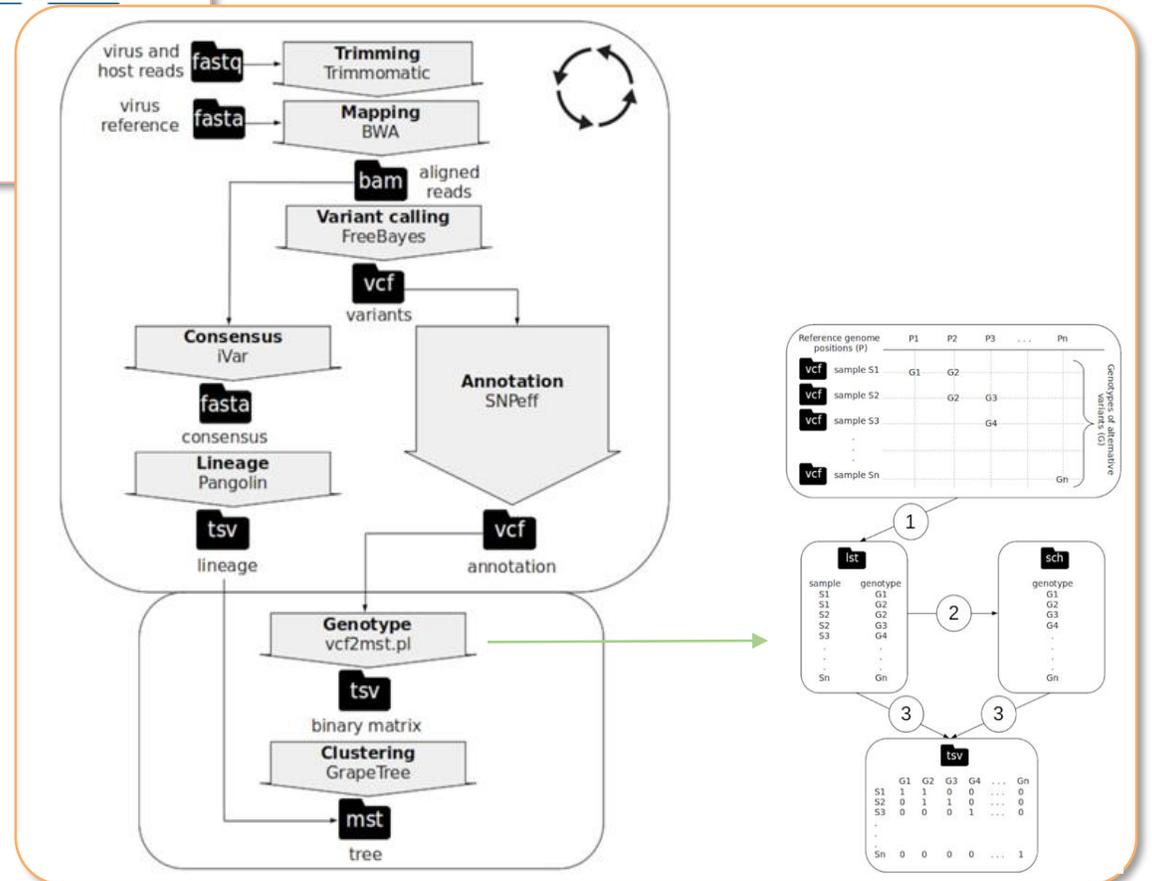
Research article | [Open Access](#) | [Published: 30 October 2021](#)

## SARS-CoV-2 surveillance in Italy through phylogenomic inferences based on Hamming distances derived from pan-SNPs, -MNP and -InDels

[Adriano Di Pasquale](#), [Nicolas Radomski](#) ✉, [Iolanda Mangone](#), [Paolo Calistri](#), [Alessio Lorusso](#) & [Cesare Cammà](#)

[BMC Genomics](#) 22, Article number: 782 (2021) | [Cite this article](#)

143 Accesses | 1 Altmetric | [Metrics](#)



**BMC Genomics (2021) 22:782,**  
**doi: <https://doi.org/10.1186/s12864-021-08112-0>**



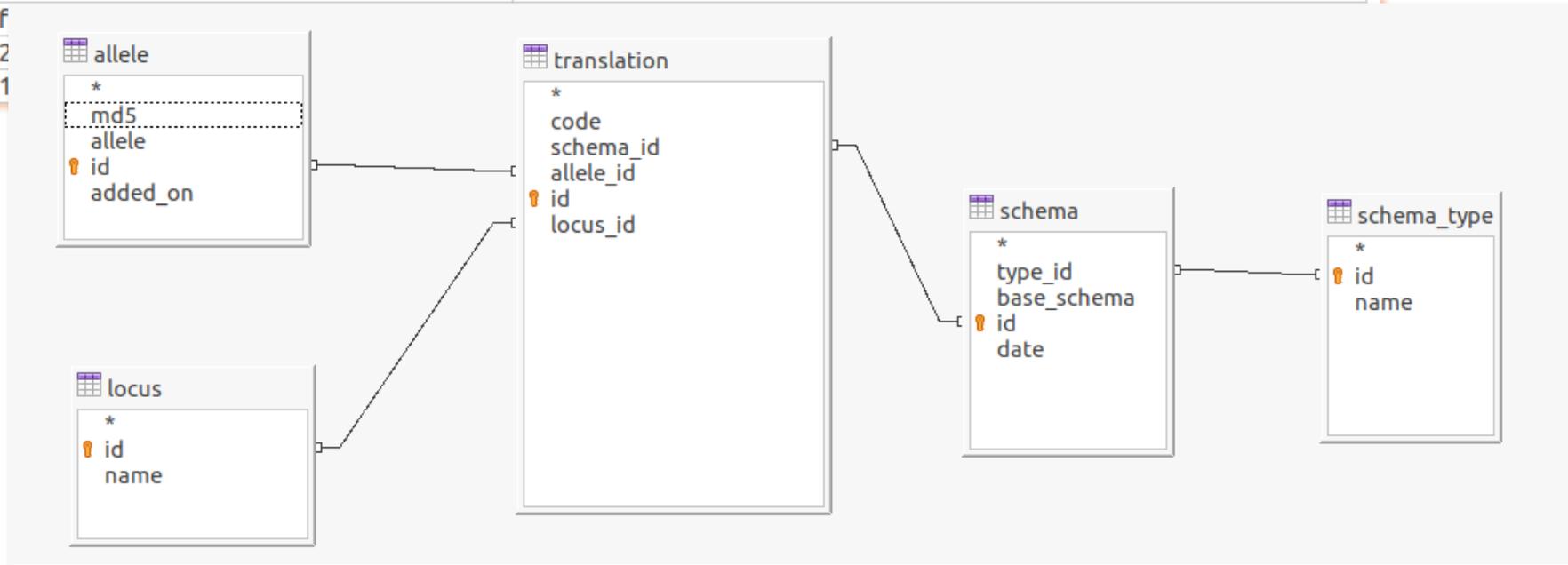
# Flexibility: INNUENDO analysis import

```
{  
  "result_patterns": [{  
    "method": "fastq",  
    "step": "0SQ_rawreads",  
    "pattern": ".*reads_num_length_num_bp_report.tab$"  
  },  
  {  
    "method": "spades",  
    "step": "2AS_denovo",  
    "pattern": ".*.*polished.fasta$"  
  },  
  {  
    "method": "mlst",  
    "step": "4TY_MLST",  
    "pattern": ".*mlst_report.txt$"  
  },  
  {  
    "method": "chewbbaca",  
    "step": "4TY_cgMLST",  
    "pattern": ".*results_alleles.tsv$"  
  }  
]  
}
```



# Analysis: Hashed based cgMLST

allele_code	locus	md5	allele
1	lmo0199	5ffc50685e04dc8ecc897df91b31d00c	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
2	lmo0199	F2273293d7505a85a43f387763cee21d	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
3	lmo0199	b1cedac79db098d4b4154c19097fce02	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
4	lmo0199	a9f18743c1638ff9a342cc7859dcd7af	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
5	lmo0199	fa6c1875b6f9bd903b81ede5118c8b17	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
6	lmo0199	f099d31c6cff96216bed006b6b42ee42	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
7	lmo0199	97c678a4059f1688b7605234ea4ae223	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
8	lmo0199	3c7d21534ee421a51da33283842baf9f	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
9	lmo0199	782f702964d85988bac610427180b629	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
10	lmo0199	fe965099842e07fc8158c109aa31e041	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
11	lmo0199	88cd6f7af1e936ebee0a9570e99b7f18	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
12	lmo0199	58e3f784f08041aa195176a805da7a24	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
13	lmo0199	1b316c4ecd58fe8a1de2da4790c7b392	ATGTCAAACGAGTATTTTGATCCAAAGTTGAAGATTTTCTCGCTAAATTCTA/
14	lmo0199	f	
15	lmo0199	2	
16	lmo0199	1	



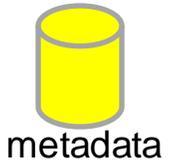
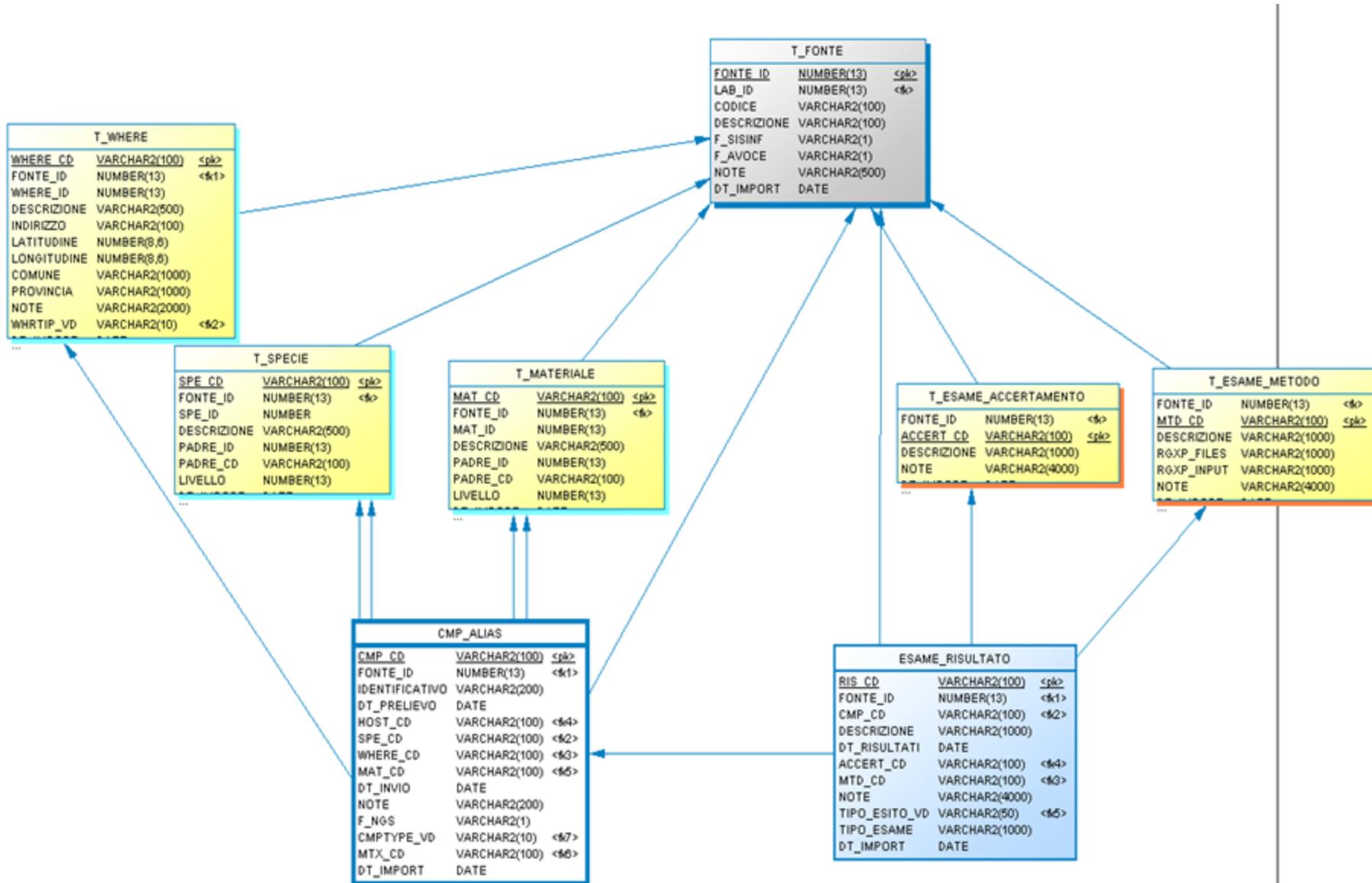


## Task 4.1. Steps of the feasibility study at MS level

- Choice of pathogen species
- List of the **involved organizations**
- List of all the useful information sources,
- A prototype system is used to let the Member State evaluate the expected improvements:
  - a) Evaluate **integration** of the systems
  - b) Evaluate the possibility to provide **EU harmonized output**.

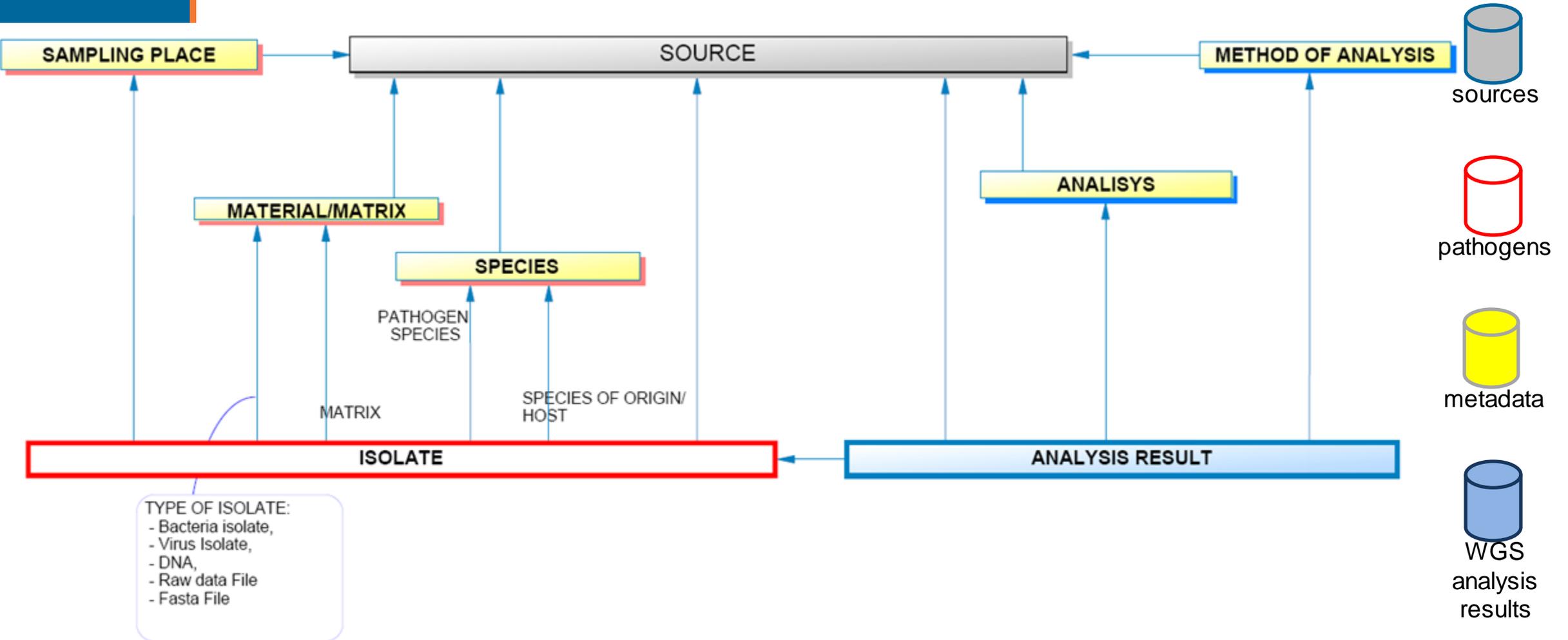


# CIS E-R diagram. Import from different sources





# CIS E-R diagram. Import from different sources



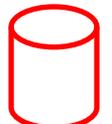


# Import different codes for the same sample

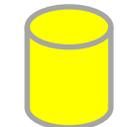
Sample	Matrix	Source
Sample1	Pizz ka a muzzarel e a pummarol ngopp	Naples Lab
Sample1	Pizza margherita	National Organization
Sample1	Pizza	EU Organization



sources



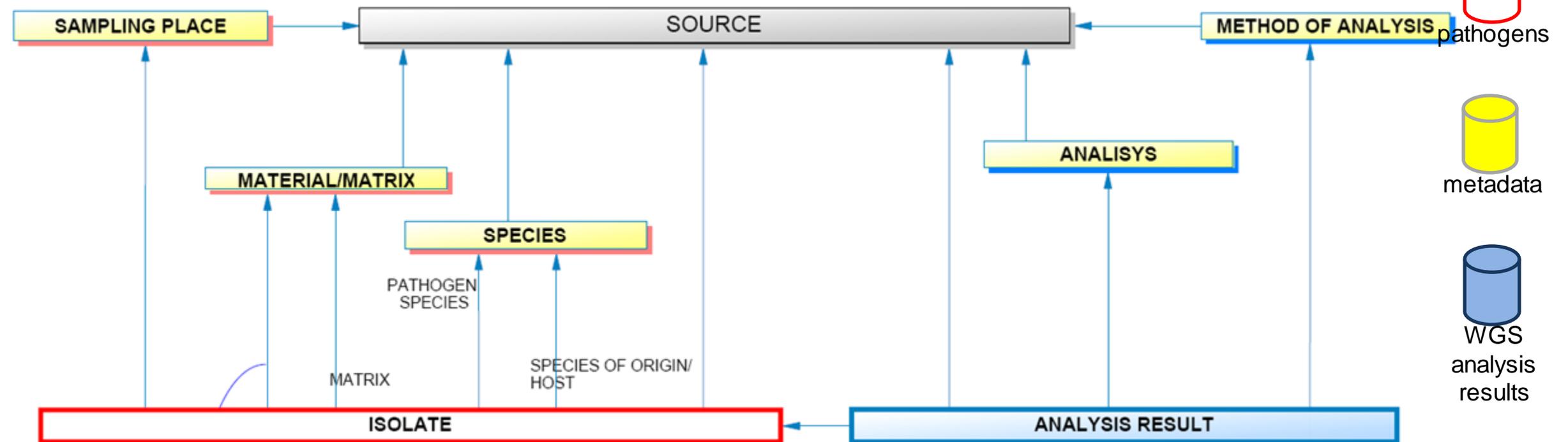
pathogens



metadata



WGS  
analysis  
results

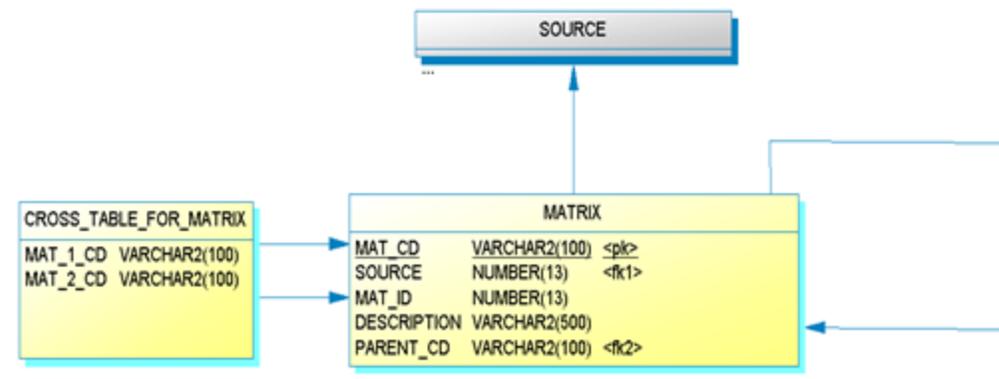




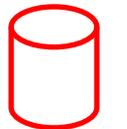
# Import different coding systems

MTX Code	Matrix	Source
Code1	Pizza ka a muzzarel e a pummarol ngopp	Naples coding system
Cod-2.1	Pizza Margherita	ISO XXX
Cod-2	Pizza	ISO XXX

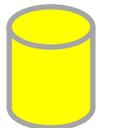
ISO XXX
<ul style="list-style-type: none"> <li>· Cod-2: Pizza               <ul style="list-style-type: none"> <li>○ Cod-2.1: Pizza Margherita</li> <li>○ Cod-2.2: Pizza Capricciosa</li> </ul> </li> </ul>



sources



pathogens



metadata



WGS  
analysis  
results



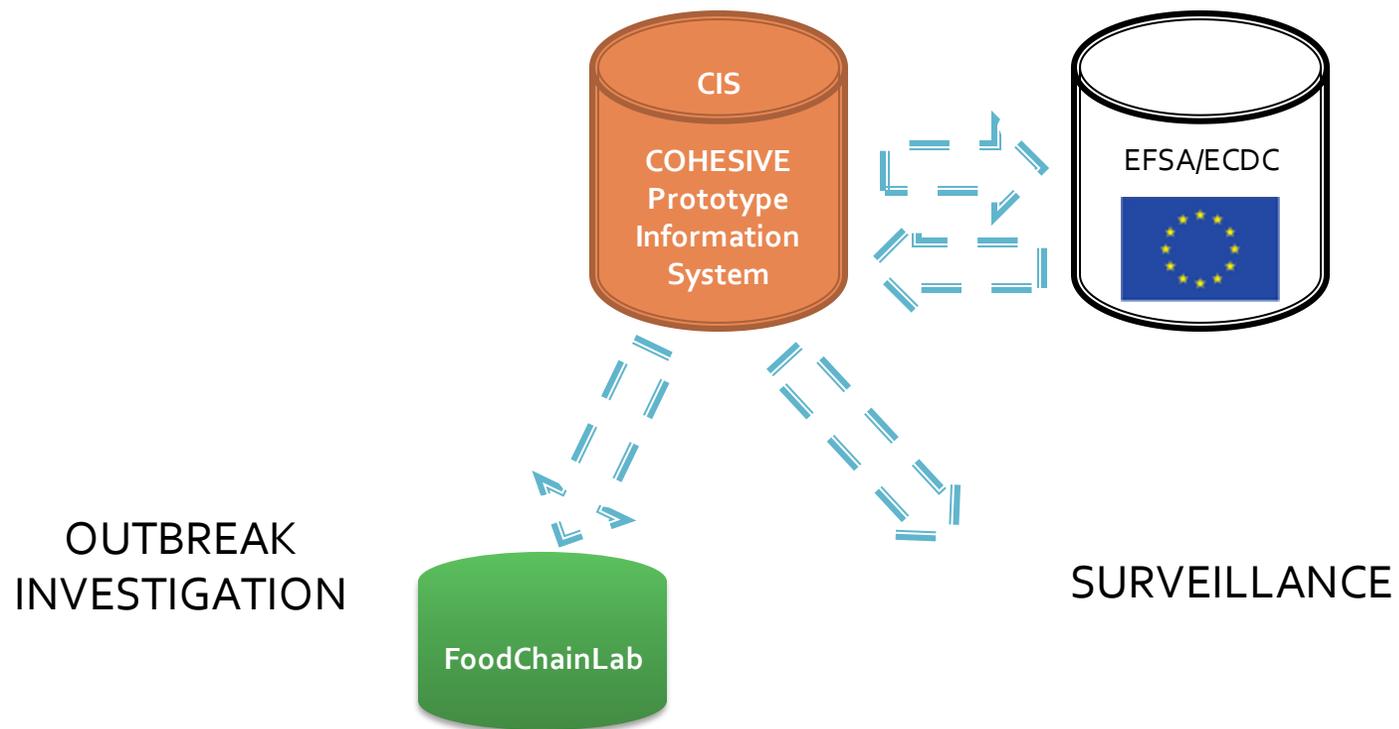
## Task 4.1. Steps of the feasibility study at MS level

1. Choice of pathogen species
1. List of the **involved organizations**
1. List of all the useful **information sources**,
1. A prototype system is used to let the Member State evaluate the expected improvements:
  - a) Evaluate **integration** of the systems
  - b) Evaluate the possibility to provide **EU harmonized output**.



# Harmonization

MTX Code	Matrix	Source
Code1	Pizza ka a muzzarel e a pummarol ngopp	Naples coding system
Cod-2.1	Pizza Margherita	ISO XXX
Cod-2	Pizza	ISO XXX





# Harmonization

Basic approach:

**Mapping between terms** → (EFSA FOODEX)

Other possible approaches:

**Ontologies** → (FOODON)



# Ontologies: a bridge between COHESIVE and ORION

EJP ORION provides frameworks for semantic data annotation using **ontologies** of relevance for surveillance data, such as FoodOn, GenEpiO and HSO (health surveillance ontology).

Data annotated manually or using automated tools (**LexMapr**) will be inserted in CIS and can promote knowledge discovery via machine learning, and interoperability across sectors.





# Ontologies: a bridge between COHESIVE and ORION

EJP ORION provides frameworks for semantic data annotation using **ontologies** of relevance for surveillance data, such as FoodOn, GenEpiO and HSO (health surveillance ontology).

Data annotated manually or using automated tools (**LexMapr**) will be inserted in CIS and can promote knowledge discovery via machine learning, and interoperability across sectors.



MTX Code	Matrix	Source
Code1	Pizz ka a muzzarel e a pummarol ngopp	Naples coding system
Cod-2.1	Pizza Margherita	ISO XXX
Cod-2	Pizza	ISO XXX

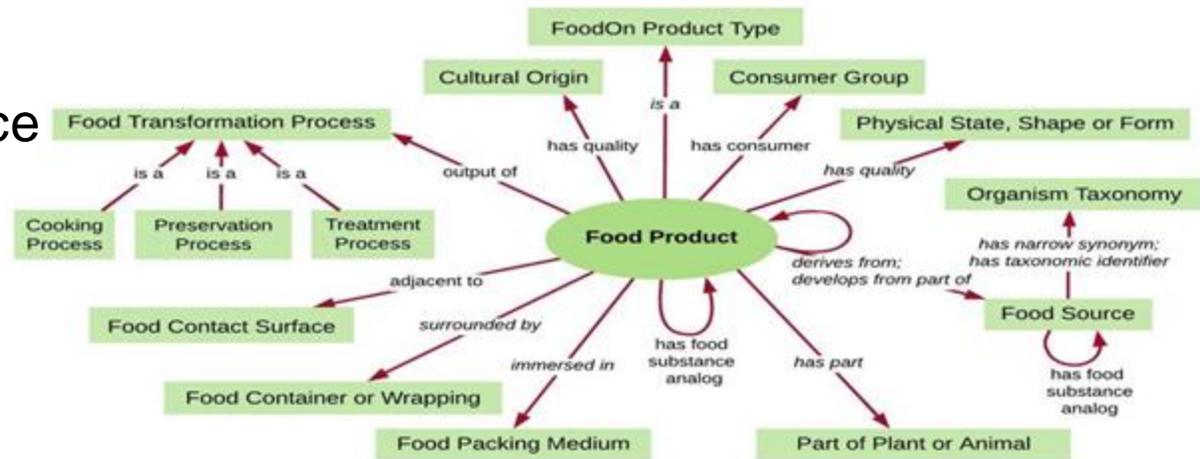




# Ontologies: a bridge between COHESIVE and ORION

EJP ORION provides frameworks for semantic data annotation using **ontologies** of relevance for surveillance data, such as FoodOn, GenEpiO and HSO (health surveillance ontology).

Data annotated manually or using automated tools (**LexMapr**) will be inserted in CIS and can promote knowledge discovery via machine learning, and interoperability across sectors.



MTX Code	Matrix	FoodOn Code	Source
Code1	Pizz ka a muzzarel e a pummarol ngopp	<u>FO-code-102</u>	Naples coding system
Cod-2.1	Pizza Margherita	<u>FO-code-102</u>	ISO XXX
Cod-2	Pizza	<u>FO-code-1</u>	ISO XXX

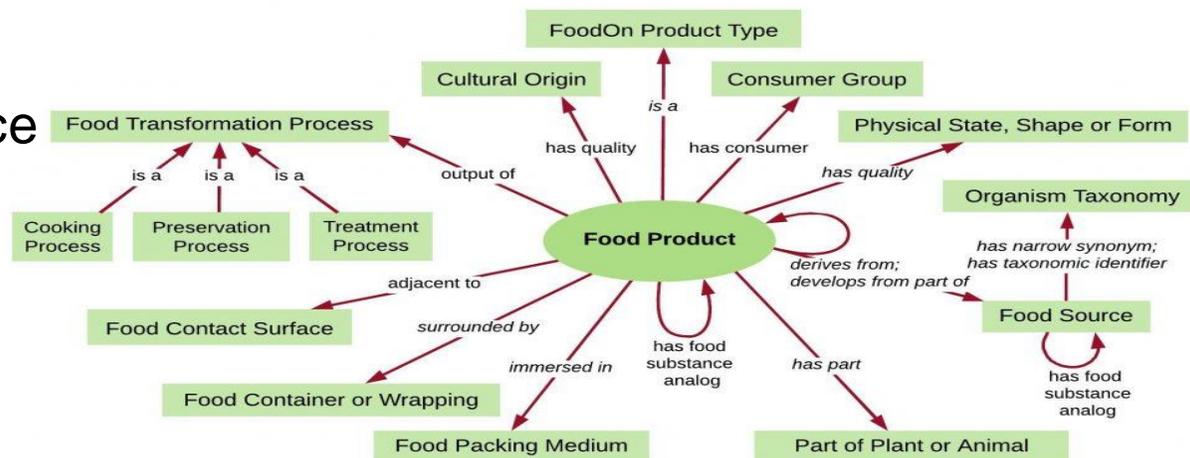




# Ontologies: a bridge between COHESIVE and ORION

EJP ORION provides frameworks for semantic data annotation using **ontologies** of relevance for surveillance data, such as FoodOn, GenEpiO and HSO (health surveillance ontology).

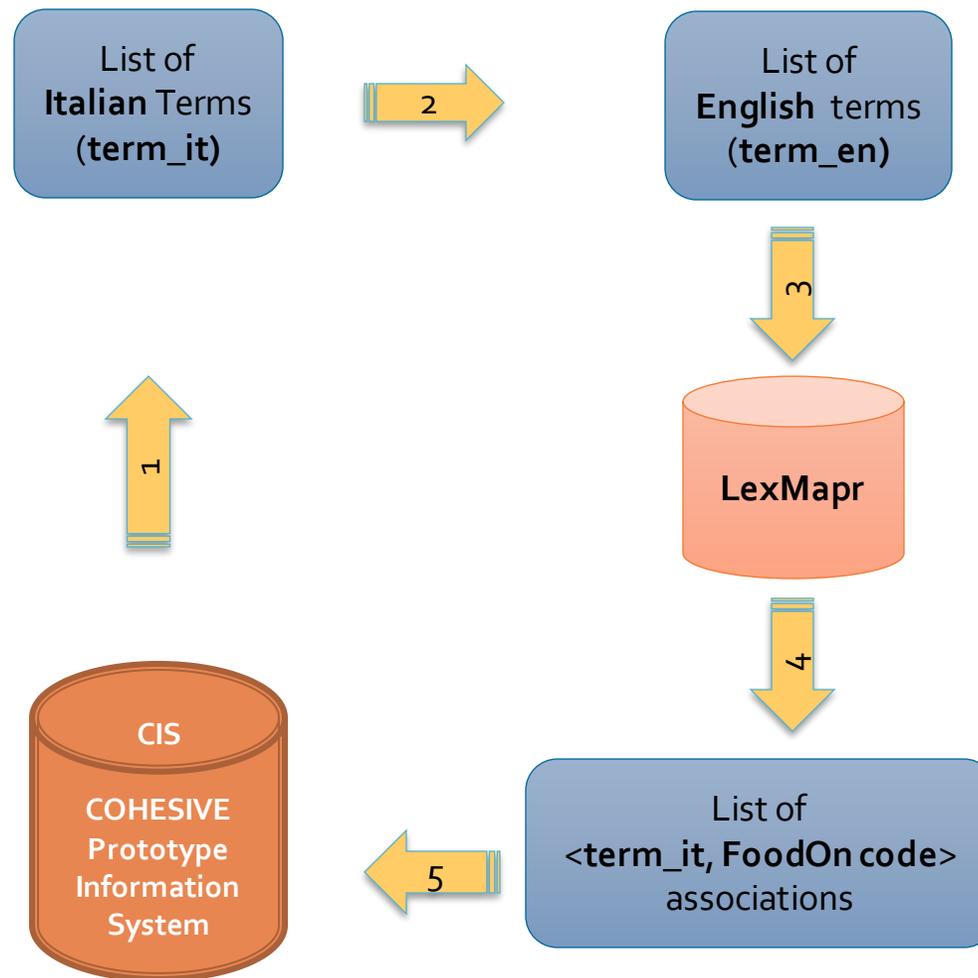
Data annotated manually or using automated tools (**LexMapr**) will be inserted in CIS and can promote knowledge discovery via machine learning, and interoperability across sectors.



MTX Code	Matrix	FoodOn Code	Source
Code1	Pizz ka a muzzarel e a pummarol ngopp	<a href="#">FO-code-102</a>	Naples coding system
Cod-2.1	Pizza Margherita	<a href="#">FO-code-102</a>	ISO XXX
Cod-2	Pizza	<a href="#">FO-code-1</a>	ISO XXX



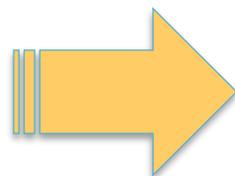
# Step1: Ontologies – Lexmapr





## Step2: Ontologies – neo4j

Using open source technologies we imported into Neo4j the FoodOn ontology in order to build the first iteration of the Knowledge Graph.



apoc-procedures (\*)  
neosemantics



(\*)

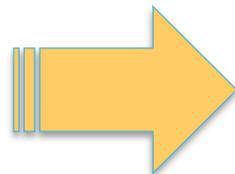
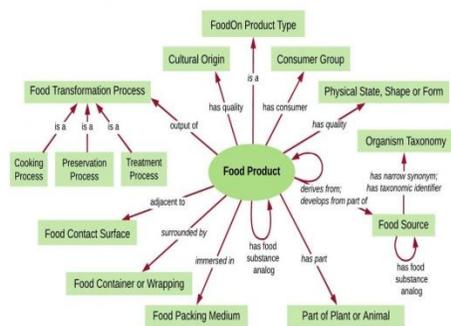
<https://github.com/neo4j-contrib/neo4j-apoc-procedures>

<https://github.com/neo4j-labs/neosemantics>



## Step2: Ontologies – neo4j

Using open source technologies we imported into Neo4j the FoodOn ontology in order to build the first iteration of the Knowledge Graph.



apoc-procedures (\*)  
neosemantics



(\*)

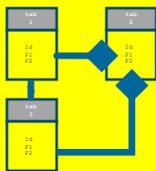
<https://github.com/neo4j-contrib/neo4j-apoc-procedures>

<https://github.com/neo4j-labs/neosemantics>

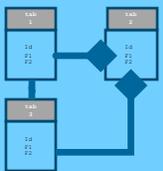


# Result: Ontologies – Query (ver. 1)

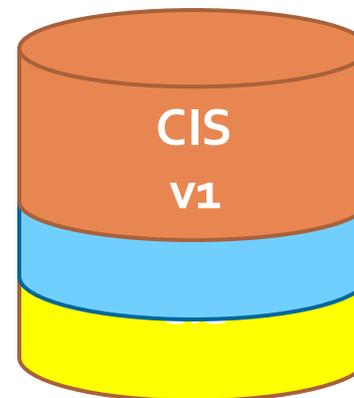
```
select samples where
```



```
sampling_date is 2018  
and  
sampling_place is Italy
```



```
and  
clonal_complex is CC8
```



WGS

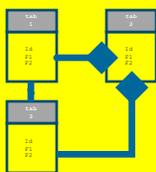


metadata

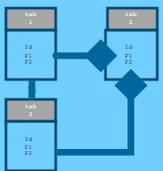


# Result: Ontologies – Query (ver. 2)

```
select samples where
```



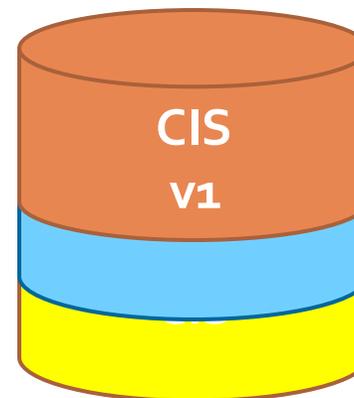
```
sampling_date is 2018  
and  
sampling_place is Italy
```



```
and  
clonal_complex is CC8
```



```
and  
samples.foodoncode  
Sub class of  
"dairy product".foodoncode
```



WGS

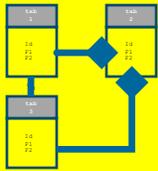


metadata

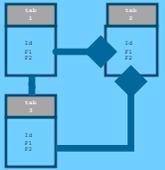


# Future Ontologies and Queries

select **samples** where



sampling\_date *is* 2018  
and  
sampling\_place *is* Italy



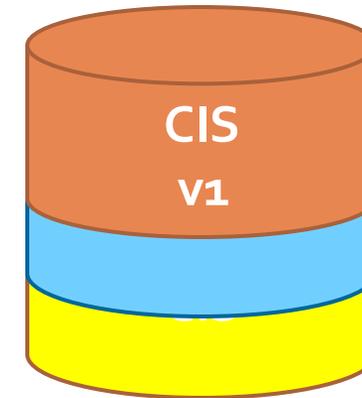
and  
clonal\_complex *is* CC8



and  
samples.foodoncode  
Sub class of  
"dairy product".foodoncode



and  
GOterm  
has relationship  
GO\_0140657:ATPase activity



WGS



metadata



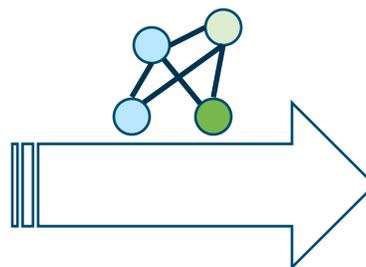
(\*)<http://www.ontobee.org/ontology/GO>



# Simple use case

Samples grouped by  
Matrix code  
in CIS

Matrix	# samples
Mozzarella (FOODON_03542735)	100
Yogurt (FOODON_03301502)	11
Beef steak (FOODON_03311663)	200
Sausages (FOODON_03542143)	22



Samples grouped by  
«Target Foodon terms»

Matrix	# samples
dairy food product (FOODON_00001256)	111
meat food product (FOODON_00001006)	222



# Status of Feasibility studies



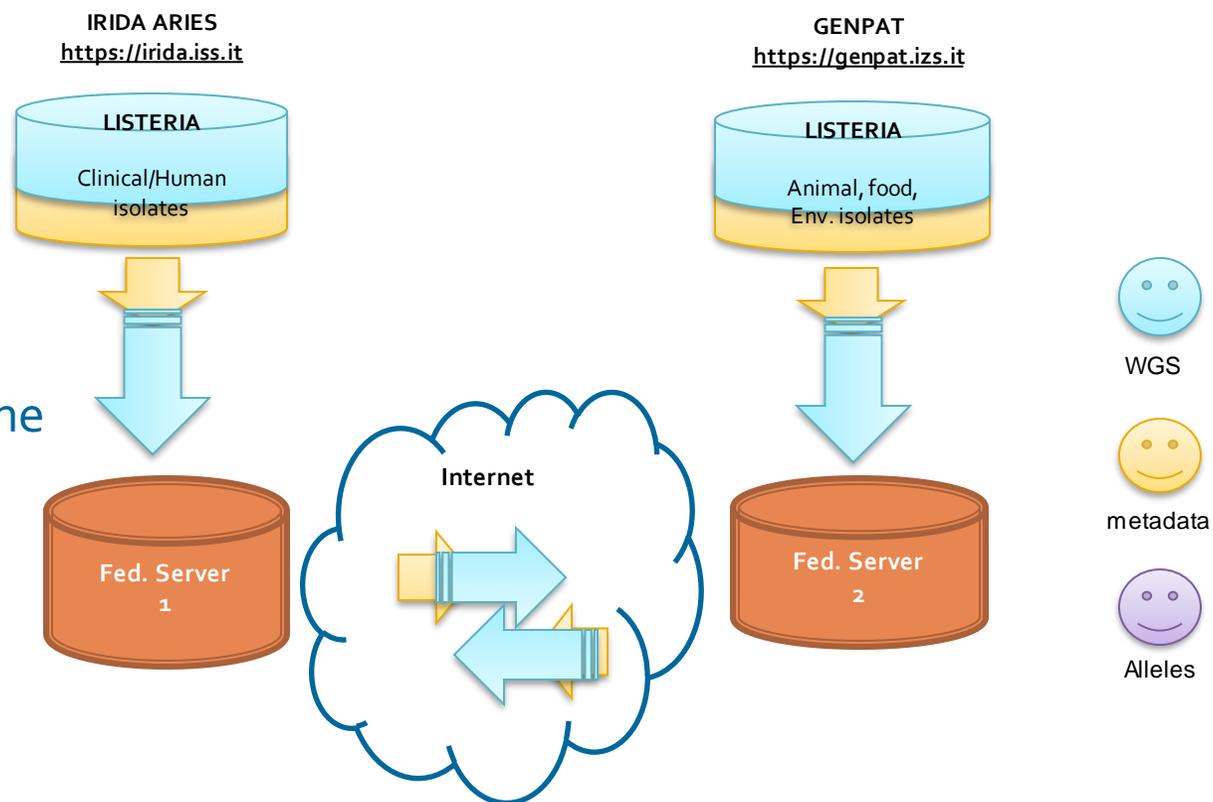
# Status of Feasibility studies

Country	Metadata Harmonization	WGS Analysis Harmonization
Italy		
The Netherlands		
Norway		



# Italian Feasibility study - Italian One-Health JointDB

Listeria based prototype of the **Italian One-Health JointDB** using federated servers





# DEMO

Demo: <https://cohesive.izs.it>

Manual: <https://cohesive.izs.it/wiki/user/>

Software: <https://github.com/orgs/genpat-it/>

# Thank you for your attention!

**Adriano Di Pasquale**  
**a.dipasquale@izs.it**



@OneHealthEJP



/company/h2020-One-Health-EJP



OneHealthEJP.eu