P02: The Galaxy framework as a concept for a national system for monitoring and surveillance of infectious disease

Arnold Knijn[1], Massimiliano Orsini[2], Valeria Michelacci[1], Stefano Morabito[1]

[1]Istituto Superiore di Sanità, Rome , Italy

[2]Istituto Zooprofilattico Sperimentale dell'Abruzzo e Molise, Teramo, Italy

Contact: arnold.knijn@iss.it

Introduction

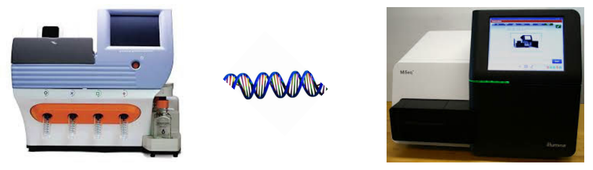
A proposal has been submitted to a national call of the Italian Ministry of health concerning the creation of a National Information System for the collection of genomic data in the field of veterinary public health, with the aim of deploying a state of the art molecular epidemiology approach to the surveillance of food-borne zoonoses and infectious diseases at the human and animal interface.

Aims

- Development of a National Information System for the collection of genomic and epidemiological data to enable the Next Generation Sequencing (NGS)-based surveillance of infectious epidemics, foodborne outbreaks and diseases at the animal-human interface.
- Development of analytical pipelines enabling harmonized, real time multi-genome comparisons, to improve the detection of clusters of cases of infections and allowing the global bio-tracing of pathogens.
- Development of metagenomics models for the culture-independent detection and typing of pathogens and the study of their interactions with the microbiota in human and animal samples and in the vehicles of infections.

Concept

NGS Data production

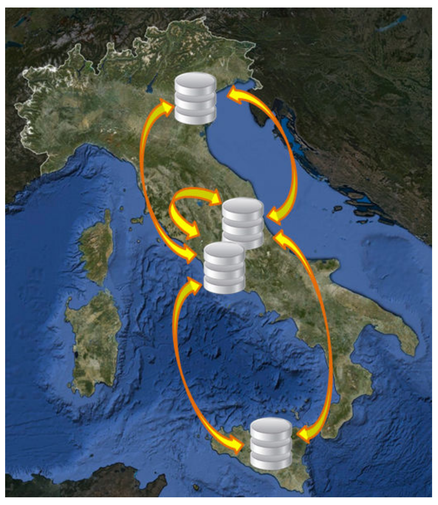


NGS Data Analysis

- Quality check
- Metagenomics
- Assembly
- ExMLST, WgMLST, WgSNPs
- Nomenclature and Ontology

NGS Data Storage

Geographically distributed database



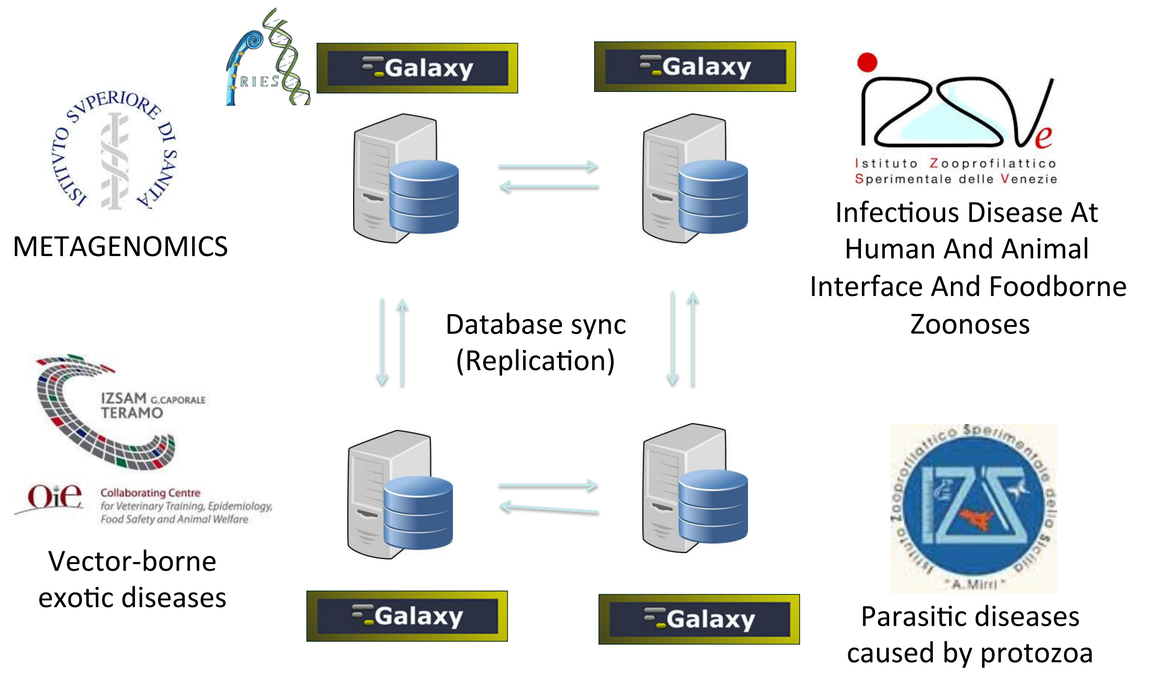
Asynchronous multi-master replication designed to provide
- High availability
- Geographically distributed disaster recovery capabilities

Each node has a local copy of the data present on all the other nodes

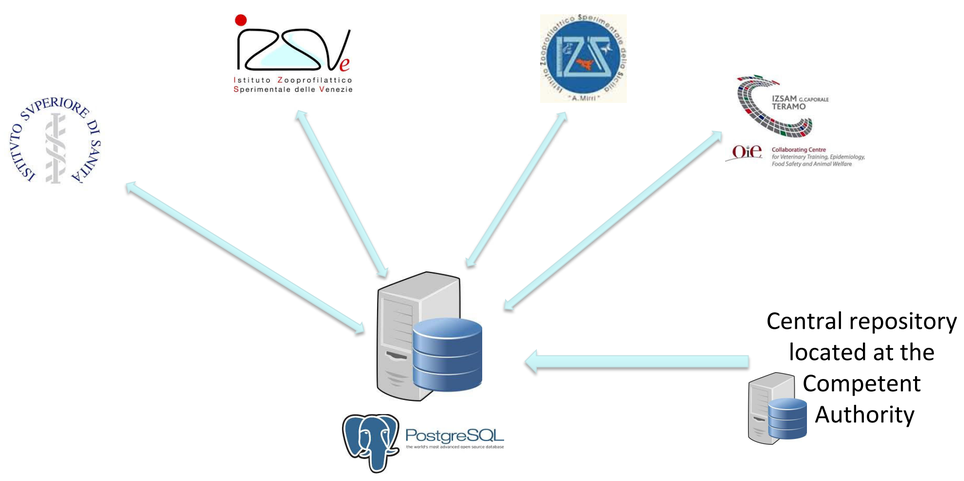
Queries run locally on individual nodes.

Each node is internally consistent at all times; the group of servers as a whole is eventually-consistent.

Network



Sensitive Data protection plan



NGS + **Unique Identifier** + a subset of non-sensitive metadata **(Accessible by external users with the credential provided by the GalaxyFW manager in agreement with the Competent Authority)**

Sensitive metadata **Not Visible by external users**

Materials and Methods

Each participating institute will produce NGS data from the pathogens concerned using the already available NGS platforms.

Repositories will be set up at the institutes coordinating the different WPs and populated with both the data produced during the project activities or available in the public databases.

A common bioinformatics environment and a web-based user interface will be discussed and developed and will be used at all the participating institutes to interact with the relational databases and to grant access to all the other participants.

The Information System used by all the nodes of the network will be implemented into the open source framework Galaxy.

Preliminary data

Facilities for NGS data production, as well as IT services are available in all the network and bioinformatics infrastructures are either under development or already established in the participating institutes.

An **Advanced Research Infrastructure for Experimentation in GenomicS (ARIES)** based on Galaxy infrastructure is already running on the servers of the Istituto Superiore di Sanità, implementing basic tools for the analysis of metagenomics data and for bacterial characterization based on whole genome sequences. The platform is planned to open for external access on January 1st 2016.

Expected results

A toolbox composed of harmonized protocols for NGS data production, manuals for the correct use of the developed bioinformatics pipelines and standard training modules will be developed and used to build up the capacity, at the national level, to produce NGS data and to interact with the developed national system for the collection of genomic and epidemiological data.

CONCLUSIONS

The development of a national system for the production, collection and analysis of genomic data on agents of foodborne zoonoses and infectious diseases will be the ground for an enhanced real time monitoring and surveillance. The harmonization within the proposed network will serve as a proof-of-concept for the whole system to be implemented at a nation-wide level, enabling the rapid and coordinated response to novel and known health threats with major benefits for the national public health.