

# Gemelli 👀

### GENERATOR

**BEYOND ONTOLOGY AWARENESS** 



**GEmelli big-data NEtwork for Retrospective Analysis and Tests in Oncology and medical Research** 

#### Carlotta Masciocchi Ph.D

Fondazione Policlinico Universitario A. Gemelli IRCCS





Centralize and integrate the logical, consistent and traversal data coming from decentralized archives such as the Datawarehouse (DWH) and the various clinical and research archives present in the institutional departments.





#### Generator









User Interface

To facilitate and automate the exploration, monitoring, collection and insertion of data already existing in the DWH or coming from external sources (Data Augmentation). The data storage must strictly follow the predefined ontology in order to guarantee the quality and usability of different types of archives.



Data Crawler



Preliminary Analysis Tools To facilitate the data entry by clinicians, allowing them to quickly insert data into the Data Augmentation archives. These robots can interact with external sources (e.g. excel file) and web application (e.g. REDCap)

to perform preliminary analysis on the data, in order to identify errors and outliers or warn in case of anomalous data combinations.







Automated Mode Creation



Internal Validation

Marketplace

To develop predictive models towards specific applications and outcomes.

To perform internal validation of the models in order to avoid overfitting (apparent improvement of model performance).

To create a functional platform to allow the use of the developed predictive models to specific stakeholders.







To create automatic reports to be used mainly for publication purposes.



External Validation To externally validate models once external data are available, in order to guarantee the reproducibility and transferability of predictive models.



Really Rapid Learning To update automatically predictive models with the addition of new data from other Institutions or based on prospective data collections.



Conclusion

Fondazione Policlinico Universitario Agostino Gemelli IRCCS Università Cattolica del Sacro Cuore

> This is a modular solution and it is applicable on vertical areas such as oncology, cardiology, neurology, geriatrics and chronic diseases, crossed with cross-sectional hematochemical, molecular, microbiota, clinical nutrition and radiology / radiomics data

• 5 proof-of-concepts (PS, Sepsi, Rectum, Breast and Ovarian cancer) are on-going to test the infrastructure and robots



## THANK YOU FOR YOUR ATTENTION