

# PE6 BOSPHORUS

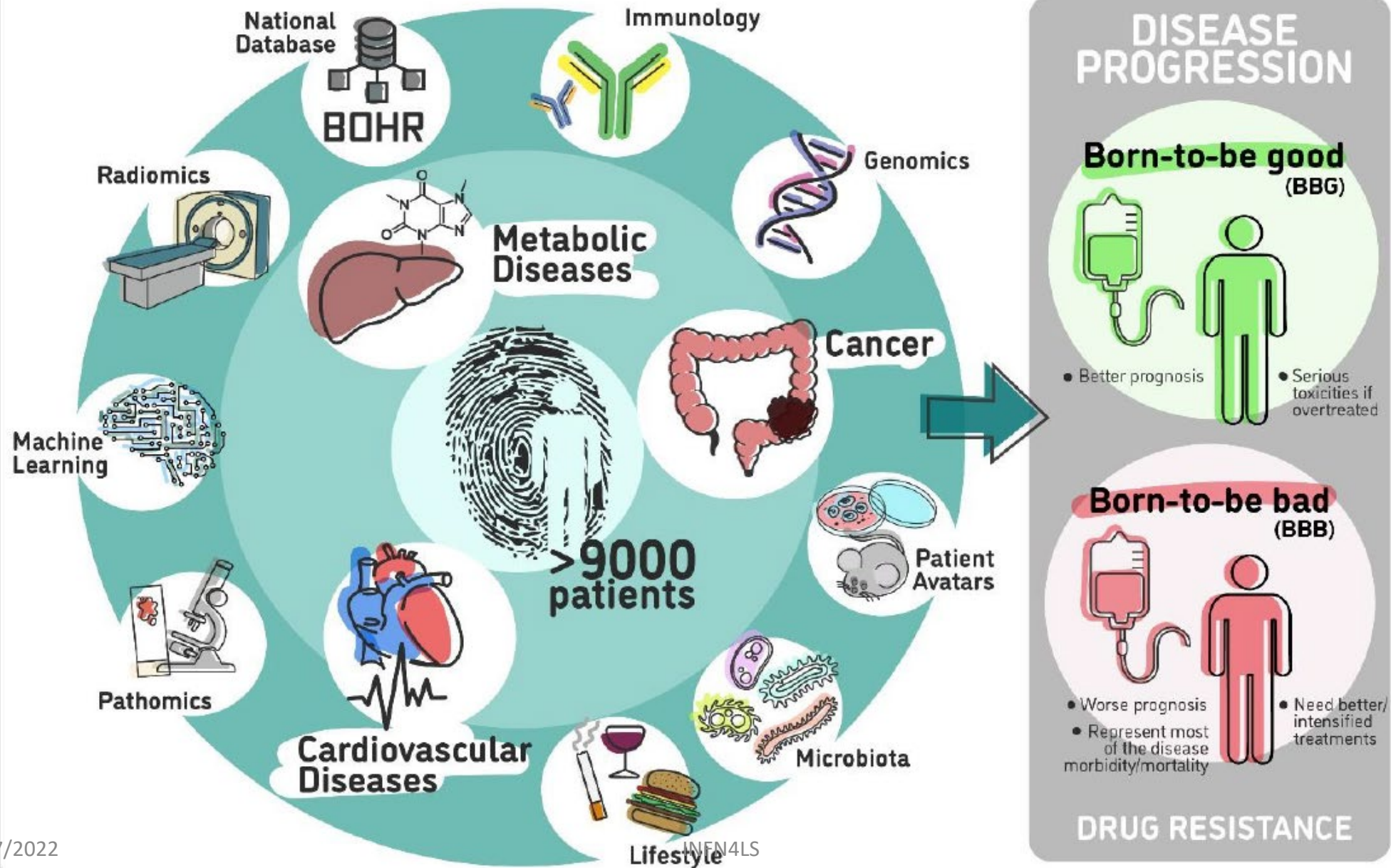
Barbara Martelli

Davide Salomoni

# PNRR Partenariato Esteso – Medicina di Precisione

- **In fase di valutazione**
- BOSPHORHUS: BOrn-to-be-good and born-to-be-bad disease: a precision medicine Survey of PHenotypes and Outcomes based on Rational interrogation of HUman Samples
- Identify the mechanisms underlying BBG versus BBB phenotypes in cancer, cardiovascular and metabolic disorders through the establishment of innovative and broadly applicable precision medicine paradigms
- Leader UniTO
- Budget totale 160 M€

# BOSPHORHUS: BOrn-to-be-good and born-to-be-bad disease: a precision medicine Survey of PHenotypes and Outcomes based on Rational interrogation of HUman Samples



# Main objectives

- generate a comprehensive patient atlas containing clinical, genomic and metagenomic information (SPOKE 1)
- identify biomarkers of disease progression to personalize risk stratification, distinguishing patients whose disease is effectively treated by standard approaches (BBG) from those who require additional and/or ideally novel treatments (BBB)(SPOKE 2, SPOKE 3 and SPOKE 4)
- optimize personalized strategies for patients with BBB pathological states who progress after standard treatments (SPOKE 2)
- create a national infrastructure for the ingestion, storage and analyses of research patient data (BOSPHORHUS National Omic-Health Research (BOHR) Athenaeum)(SPOKE 5)
- leverage omics and clinical data to generate mechanistic hypotheses on the molecular bases of cancer progression and resistance to therapy, which will be tested in innovative preclinical models (SPOKE 6)



# Governance – the hub

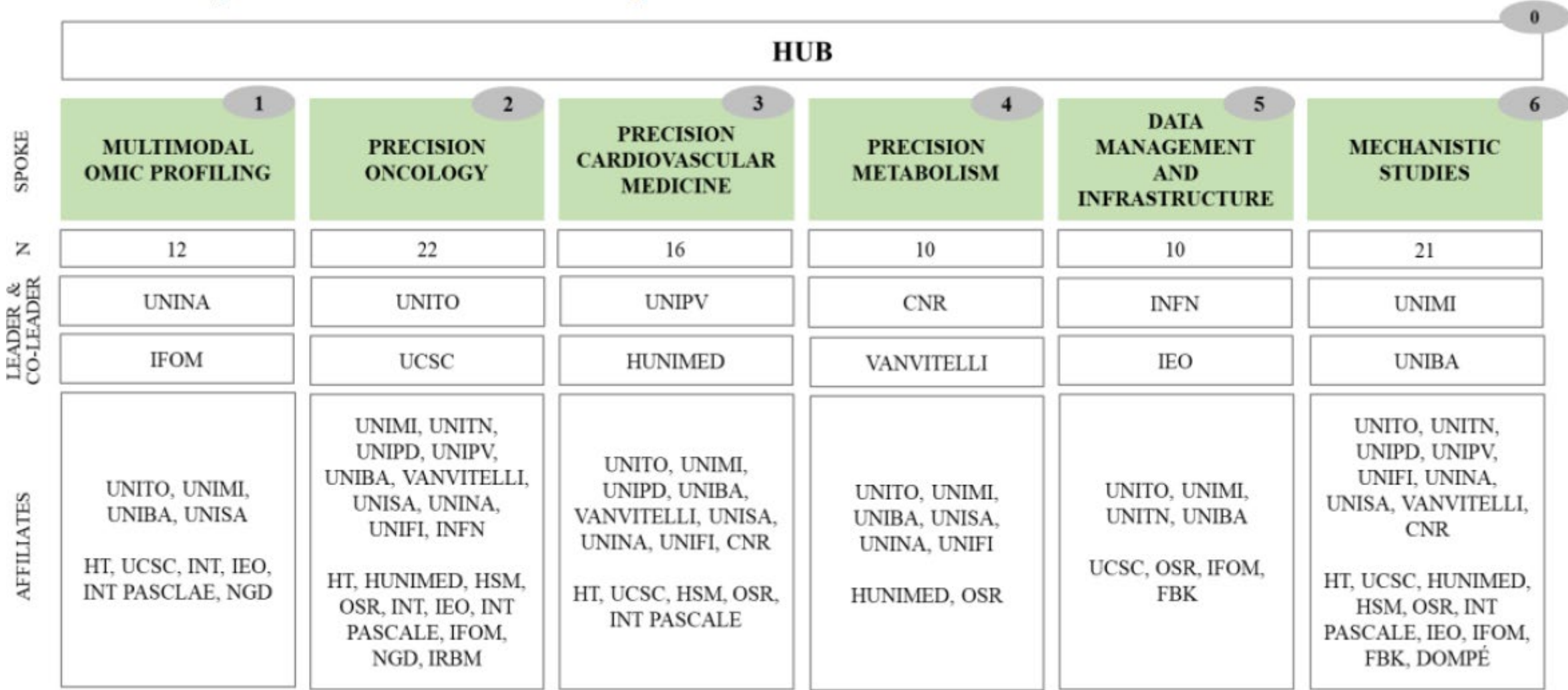
- The Hub will have the legal form of a S.c.a.r.l and it will be located in Sicily
- management and administrative accounting control on its expenses and on those that are incurred by the Spokes
- responsible for the EP and supervise the correct implementation of the work plan and its compliance with the milestones and timeline
- will ensure the accomplishment of the macro-objectives collaborating with the Spoke Leaders



# Governance – the spokes

- 6 thematic spokes organized and regulated by agreements among the parties
- Spoke Leader tasks:
  - carrying out the project activities and reporting to the Hub
  - manage the relationships with the Spoke Affiliates
  - realize a synergistic interconnection with the territory, by exploiting the whole range of available measures, including the Open Calls
  - overall management and coordination of the “Spoke Affiliates
  - monitoring the scientific and technical progress of the entire Spoke, the knowledge management, other innovation related activities
  - manage the relationship with the Hub

# BOSPHORHUS EP architecture





# Spoke relationship

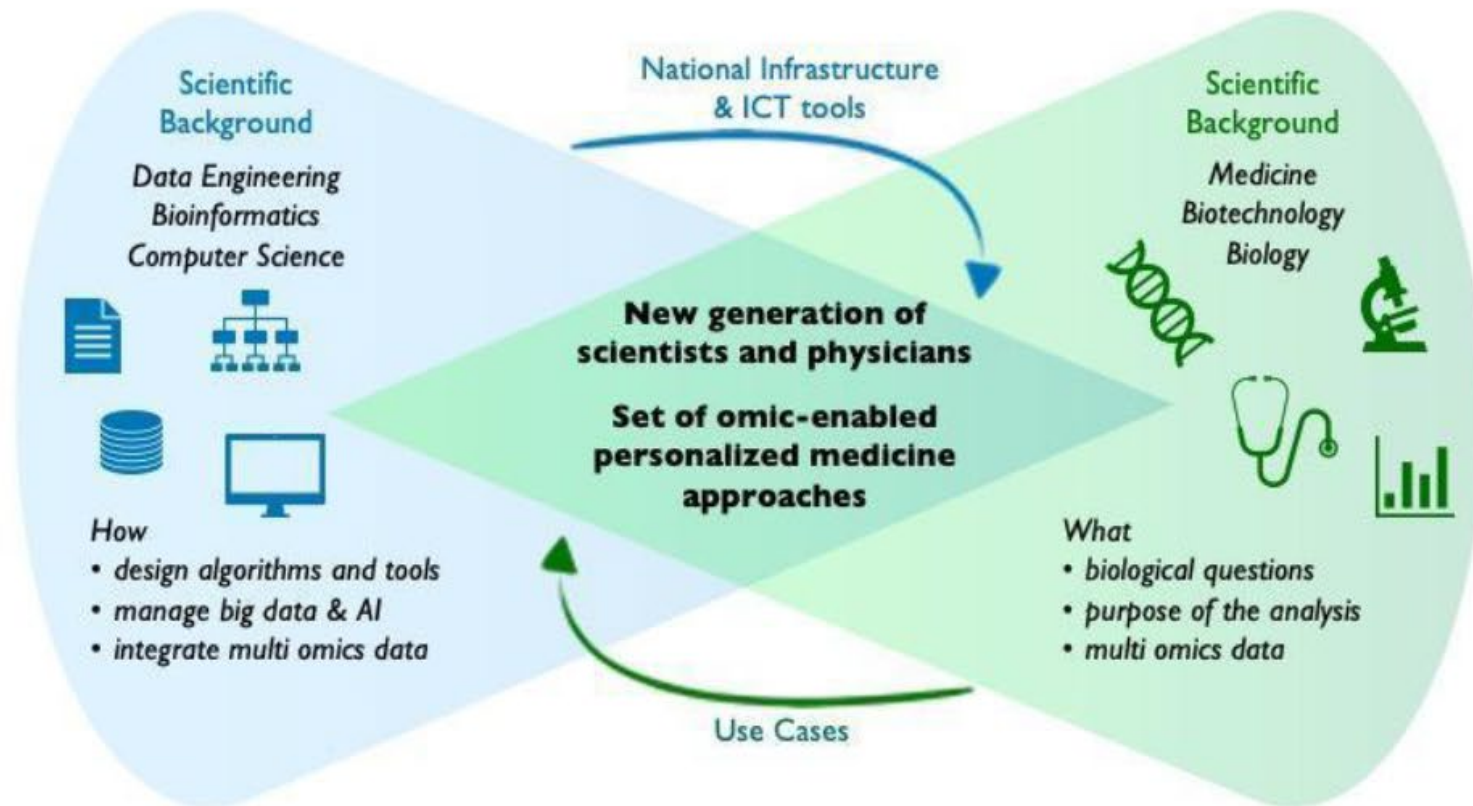
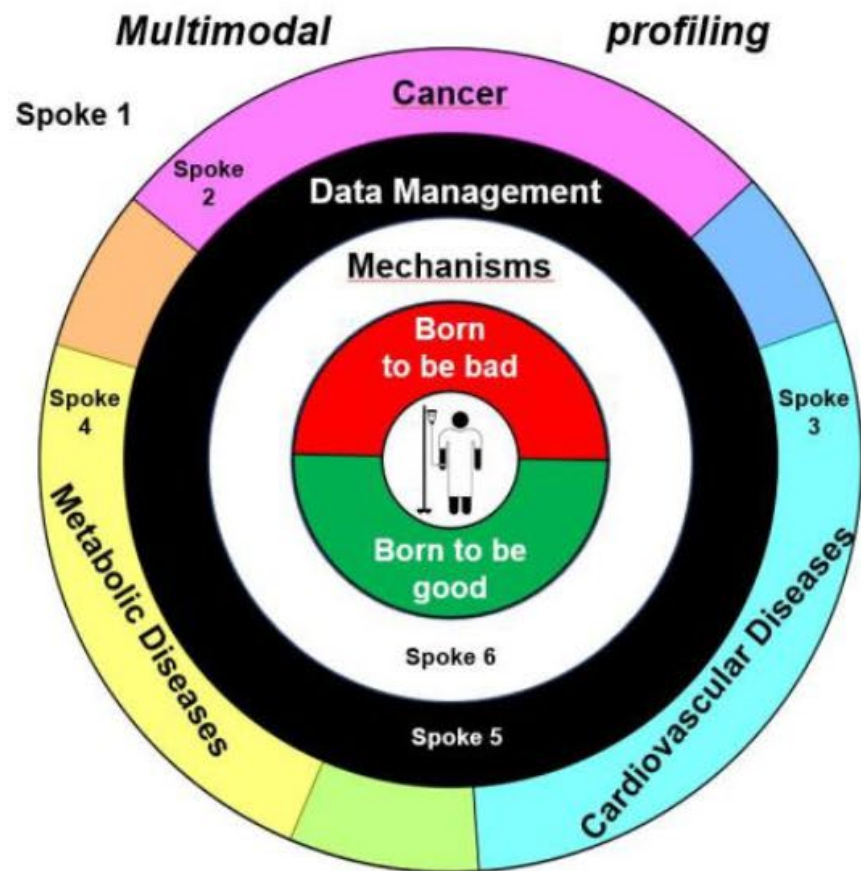


Figure B.13: Graphical representation of BOSPHORHUS Spoke interplay (left) and highlights on data analysis (right)

# INFN contribution

# Affiliate in Spoke 2

- Leader: UNITO – Alberto Bardelli, Co-leader: UCSC – Ruggero De Maria
- Main objective: Create and infrastructure a multidisciplinary national community focused on science-intensive markers discovery to **predict cancer progression and treatments resistance**, and develop discerning care trajectories for BBG vs. BBB malignancies
- Spoke 2 is the clinical powerhouse of BOSPHORHUS for the cancer space. **7,000 cancer patients will be enrolled in the BOSPHORHUS master observational trial (BOHR-MOT)**, an innovative trial construct scaled up to a **pan-cancer** space for the first time in Italy.
- Patients will be profiled at unprecedented systemic and tumor granularity anchoring multi-omics results, including radiopathomics, to a genomic backbone

# Spoke 5 – Data Management and Infrastructure

- Leader INFN, Co-leader IEO
- Affiliates: UNITN, UNIMI, UNITO, UNIBA, UCSC, IFOM, OSR, FBK
- Main objective: **Creation of a national platform, database and related tools for the ingestion, management and analyses of research patient data** (BOSPHORHUS National Omic-Health Research (BOHR) Athenaeum)
- BOSPHORHUS will generate the largest collection of clinically-annotated multi-omics patient data in Italy. Patient's data will be collected within a National Database and Platform -BOHR, that will:
  - function in tight connections with the National Health System to optimize clinical applications of omics technologies and improve personalized prevention, diagnoses, and treatments
  - provide health data and advanced computational expertise and tools, including AI-based approaches, to accelerate advanced biomedical research
  - be accessible to the national/international scientific community, pharma and biotechs, charities and governmental/regional departments

# Spoke 5: the big data platform

- In order to ingest, store, manage and analyze this corpus of data, a **national, open and federated Cloud and HPC platform** will be built, in close cooperation with existing and future national infrastructures for life science data
- The platform will cover a continuum of use cases to **seamlessly connect edges to cloud or HPC resources**, supporting all the Spokes of the project. It will focus on four integration scenarios:
  - 1. central harvesting of data collected remotely
  - 2. edge anonymization, followed by central ingestion and analysis of data
  - 3. edge feature extraction, followed by central ingestion and analysis of features
  - 4. federated machine learning based on edge-based training, followed by publishing of the trained methods and by inference performed either centrally or at other edge locations

# Spoke 5 – WP 1

- **Development of a plan to guarantee data sharing and access, including legal issues (General Data Protection Regulation - GDPR)**
- **Lead: INFN**; Parts.: FBK, IEO, UNIMI, IFOM.
- T1.1: Analysis of the state of the art for data sharing and access. This task starts from similar activities ongoing in projects such as Health Big Data, EOSC-Pillar, DICE, ELIXIR, Harmony Plus; it includes the definition of the Legal Framework (GDPR, IPR, FAIR principle, OA and OS), a gap analysis (definition of the most important legal gaps to data sharing) and an analysis of best practices and Guidelines, also in the EOSC context. This task will take into key account the need to harmonize multiple processes and procedures defined by the BOSPHORHUS actors and partners.
- T1.2: Elaboration of a Data Usage Agreement and of a project-wide Data Management Plan (DMP) for the collection, analyses, re-analyses and sharing of personal data, based on patient informed consent and other legal basis at national and EU level.
- T1.3: Revision and continuous monitoring of all documentation on patient personal data and submitted for ongoing research projects.
- T1.4: Evaluation of privacy and legal aspects involving patient & sample re-analyses in research projects.



# Spoke 5 – WP 2

- **Creation of automated pipelines for data collection, extraction and management**
- Lead: OSR; Parts.: FBK, INFN, UNITO

# Spoke 5 – WP 3

- **Generation of computational workflows for advanced multi-omics data analytics**
- Lead: UNITN; Parts.: FBK, IEO, UNITO, OSR, UNIBA, UNIFI, HUNIMED, HSM.

# Spoke 5 – WP 4

- **Data platform, infrastructural testbeds and adaptation of data analysis to the BOHR Athenaeum platform**
- **Lead: INFN**; Parts.: OSR, IEO, FBK, UNITO, IFOM, UNIMI, CNR
- **The BOHR Athenaeum database will be developed on a computing and data platform at INFN premises.** An initial version of the platform is already deployed at INFN, as part of the Health Big Data Project. The activities of WP4 will be performed in strong collaboration with all the other WPs of this Spoke and with the other Spokes of the project, so that specific analysis workflows and pipelines developed in those contexts can be executed on the integrated platform, and in close connection with ELIXIR activities, leveraging community experience in using and developing state of the art DM solutions (e.g. RDMkit)
- T4.1: define the BOHR Athenaeum multi-source database structure. The database will include both structured and unstructured data, coming from multiple sources, and will scale horizontally
- T4.2: define an integrated centralized infrastructure and an ad-hoc cloud environment to ensure: a) patient-data security and de-identification, compliant with legal obligations - GDPR; b) proper collection and storage of health data aligned with FAIR principles; c) access to the scientific community; d) federability and interoperability with the National Health System, relevant national and international RIs, and ongoing international Personalized Medicine initiatives
- T4.3: create a series of digital platforms to ensure access to bioinformatic pipelines and computational platforms, integrating data collected by the patient-reported digital infrastructure provided in WP5 on outcomes and health data coming from wearable sensors
- T4.4: create infrastructural testbeds at both the edge and cloud levels to be used by the project and set up the integrated data platform that will be used by the entire project to analyze the data
- T4.5: integrate an existing federated web system (“Cohort Genomic Platform”) to assist clinicians and researchers in the collection, storage and sharing of medical and family history, genetic information and biological samples from well-characterized patient cohorts

# Spoke 5 – WP 5

- **Interoperability with the structuring NHS database (Fascicolo Sanitario Elettronico) and with existing patient-reported digital infrastructures**
- Lead: FBK; Parts.: OSR

# Spoke 5 – WP 6

- **Training and validation of query algorithms for biomedical research**
- Lead: OSR; Parts.: FBK, IEO

# Spoke 5 – WP 7

- **AI-based solutions for diagnostics and therapy**
- **Lead: INFN**; Parts.: UNIBA, FBK, OSR, UNITO, UNIMI, IEO
- Artificial Intelligence in Medicine (<https://www.pi.infn.it/aim/>) is an ongoing research project using AI-based solutions, including radiomics, machine learning and deep-learning approaches, to derive quantitative information from data, and to build predictive models in medical diagnostics and therapy. WP7 will fully use the data sets collected by BOSPHORHUS and, strongly linking with Spokes 1 and 2, will cover two of the main challenges in this field
- T7.1: handling data samples of limited size to build AI models with reliable and reproducible performance
- T7.2: making these AI models trustworthy. A key goal to achieve this is to make AI models and their responses understandable by humans. Explaining why an AI-based system has provided a specific response on a given input is extremely important in the field of medical applications, and deserves specific technical developments, implementation, and large-scale validation



# Spoke 5 – WP 8

- **Federated Learning, Architecture and methods for privacy-preserving AI**
- Lead: UNITO; Parts.: INFN, OSR

# Spoke 5 – WP 9

- **Activation of the ecosystem of BOHR stakeholders**
- Lead: IEO; Parts.: UNIBA, UNIMI, UNIPV, INT PASCALE, UNITO, INFN
- develop a strategic plan to promote the usage of BOHR research-data by different stakeholders to foster academic, industrial and societal research, and comply with the PNR guidelines of the National Plan for Open Science

# Spoke 5 – WP 10

- **(Gen)omics and Society**

- Lead: UNIMI; Parts.: UNIBA, UNITN, OSR, INFN

- A forum of experts and patients will be assembled to promote the public debate on health data security, ethics, legal frameworks, and usage, to build continuous trust and transparency around the program (Genomics & Society Board)

# Budget overview in M€

	Permanent Staff	New RTDa	PhD	Other types of contracts	Overheads (15% of the Personnel cost)	Personnel	Materials & Licences	Consulting services	Others	Hub Services & Management	TOTAL
<b>HUB</b>										5.500.000	5.500.000
<b>SPOKE 1</b>	2.697.750	2.700.000	1.125.000	600.000	1.068.413	8.191.163	3.664.107	258.000	16.448.800		28.562.070
<b>SPOKE 2</b>	6.936.750	4.930.000	1.575.000	1.941.250	2.307.450	17.690.450	6.294.804	6.761.950	19.848.998		50.596.202
<b>SPOKE 3</b>	2.416.500	2.550.000	1.275.000	790.000	1.054.725	8.086.225	2.282.650	548.149	6.515.000		17.432.024
<b>SPOKE 4</b>	2.875.500	3.000.000	750.000	675.000	1.095.075	8.395.575	2.221.106	4.648.539	950.000		16.215.220
<b>SPOKE 5</b>	1.937.250	1.650.000	675.000	1.205.000	820.088	6.287.338	567.513	2.040.000	0		8.894.850
<b>SPOKE 6</b>	5.711.625	5.420.000	2.100.000	1.960.000	2.278.744	17.470.369	9.245.945	5.113.315	970.000		32.799.629
<b>TOTAL</b>	22.575.375	20.250.000	7.500.000	7.171.250	8.624.494	66.121.119	24.276.125	19.369.952	44.732.798	5.500.000	159.999.993

# Personale

- Spoke 2 Precision Oncology
  - Cuttone (Catania)
  - Fiorina (Torino)
  - Retico (Pisa)
- Spoke 5 Data Management and Infrastructure
  - Salomoni (spoke leader)
  - Martelli, Duma, Ronchieri (CNAF)
  - Donvito, Antonacci (Bari)
- Referenti amministrativi
  - Allegro, Haghshenas (Bologna)

# Budget INFN

	Sud	Nord	Totale
Spoke 2	129375	129375	258750
Spoke 5	130237	785625	915862
Rendicontazioni personale	185625	297000	482625
<b>TOT</b>	<b>445237</b>	<b>1212000</b>	<b>1657237</b>

Da aggiungere:

- Budget per infrastruttura circa 2M€
- Assunzione di 1 TD e 1 PhD al sud
- Assunzione di 1 TD e 1 PhD al nord
- Overhead (15% del costo del personale), circa 72 k€