



Ministero dello Sviluppo Economico

DM 31 DICEMBRE 2021 – ACCORDI PER L'INNOVAZIONE



Istituto Nazionale di Fisica Nucleare

INFN-LNS contribution to the RESILIO project

INFN4LS meeting

INFN-LNS, 19 Luglio 2022

Dr G A Pablo Cirrone, Dr Alberto Sciuto



Tecnologie digitali fondamentali, Intelligenza Artificiale, Internet delle
Cose e *Quantum Machine Learning* per la Resilienza Ambientale

Summary

- Project Resilio
 - Prospetto generale del progetto e dell'OR4 (INFN-LNS)
- INFN-LNS Contribution and Quantum Machine Learning
 - Contributo al progetto e generalità sulle tecniche utilizzate

Project Resilio - Obiettivi



– creazione di una **Environmental Intelligent Platform**, catalogabile come uno strumento decisionale basato sui dati (**data-driven decision framework**).

La nozione di Environmental intelligent platform include al proprio interno diverse tecnologie, astrazioni teoriche e applicazioni di elevato livello di complessità.

Un approccio sintetico alla progettazione della piattaforma e delle sue specifiche componenti impone una **sistematica revisione degli attuali strumenti e tecnologie di Artificial Intelligence (AI), di cybersecurity, di Big Data Analytics (BDA)** nel contesto, complesso per sua stessa natura, degli ecosistemi ambientali e delle relative interazioni.

Project Resilio - Obiettivi



- Messa a punto di un **modello spaziale e predittivo** basato sull'analisi di parametri ambientali, meteorologici, climatici, della qualità dell'aria e delle fonti energetiche rinnovabili basato sull'utilizzo di algoritmi avanzati di AI e Machine Learning (ML).
- sviluppo di un **metodo analitico per riprodurre la crescita di tessuto cellulare** partendo dalla singola cellula danneggiata, applicando algoritmi di Quantum Machine Learning (QML) con l'obiettivo di coprire il gap oggi esistente nella capacità di collegare il danno al micro/nano-metrico al DNA e il successivo sviluppo di una patologia macroscopica in funzione dell'evoluzione dei dati ambientali.

Project Resillo



La realizzazione di **nuovi sistemi elettronici volti all'acquisizione di alcuni parametri ambientali** legati alla qualità dell'aria di piccole dimensioni, a basso consumo, in grado di trasmettere informazioni automaticamente e in tempo reale, adattabili a diverse tipologie di sensori e basati su protocolli IoT allo stato dell'arte, anche in termini di cyber sicurezza.

La realizzazione di una **control room che si integrerà con l'ecosistema ICT di ARPA Sicilia** e che farà uso di tecnologie allo stato dell'arte, ospitando gli ambienti di test e produzione Resillo.

Project Resillo



Il progetto Resillo è strutturato:

In un primo gruppo di attività di **Ricerca Industriale**, volte a identificare, normalizzare, aggregare, strutturare, correlare e georeferenziare i dati ambientali affinché possano essere utilizzati come input in avanzati algoritmi predittivi, basati anche **sull'intelligenza artificiale e sul "quantum machine learning"**.

Da un gruppo di attività di **Sviluppo Sperimentale** volte alla realizzazione di un **Progetto Pilota** nella forma di una innovativa e tecnologica Control Room presso i locali recentemente ristrutturati da ARPA Sicilia.

Project Resillo

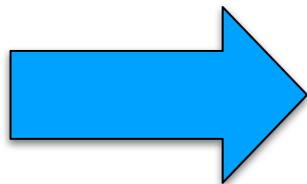
Le attività del progetto **Resillo** sono suddivise in 8 **Obiettivi Realizzativi**:



OR	Soggetto proponente	Titolo OR	RI/SS
OR1	Arpa Sicilia	Modello integrazione e correlazione dei dati, Governance Ambientale	RI
OR2	Corvallis Srl	Analisi e progettazione <i>environmental intelligence platform</i>	RI
OR3	Tor Vergata	Modelli spaziali e predittivi di parametri ambientali	RI
OR4	INFN	<i>Quantum Machine learning for data analysis</i>	RI
OR5	Infobiotech Srl	IoT sensors network	RI
OR6	Arpa Sicilia	Infrastruttura dati e Control Room	SS
OR7	Corvallis Srl	Sviluppo <i>environmental intelligence platform</i>	SS
OR8	Infobiotech Srl	Data Hub	SS

Project Resillo

Le attività del progetto **Resillo** verranno avviate a **gennaio 2023** Per terminare a **dicembre 2025**



Attività			Partner	SS/RI	Sem1	Sem2	Sem3	Sem4	Sem5	Sem6	
OR1: Modello integrazione e correlazione dei dati, Governance Ambientale											
Attività 1.1	Analisi delle sorgenti dati, delle piattaforme tecnologiche e dei middleware di integrazione (Data Hub)		ARPA	RI							
Attività 1.2	Definizione dei TAG e del Data Catalog										
Attività 1.3	Analisi e Definizione del Modello di Correlazione Geografico										
Attività 1.4	Analisi e Classificazione dei Riferimenti Normativi e Modello di Controllo Ambientale										
Attività 1.5	Analisi e Definizione del Modello di Governance Ambientale										
OR2: Analisi e progettazione environmental intelligence platform											
Attività 2.1	Analisi dello stato dell'arte delle tecnologie, dei requisiti di sistema e specifiche tecnico funzionali		CORV	RI							
Attività 2.2	Analisi e progettazione algoritmi di scientific advanced analytics e Artificial intelligence										
Attività 2.3	Studio di una soluzione per la sicurezza della comunicazione ed analisi del traffico dei dispositivi IoT										
Attività 2.4	Progettazione architettura Environmental intelligence platform										
Attività 2.5	Design interfaccia utente Environmental intelligence platform (webapp, webGIS, dashboard)										
OR3: Modelli spaziali e predittivi di parametri ambientali											
Attività 3.1	Identificazione ed analisi dei dati ARPA Sicilia e dei dati disponibili presenti nei servizi nazionali, europei e internazionali.		TORV	RI							
Attività 3.2	Identificazione delle variabili ambientali che dovranno essere visualizzate e studiate.										
Attività 3.3	Identificazione, sviluppo e validazione degli algoritmi di spazializzazione e predizione delle variabili ambientali.										
OR4: Quantum Machine learning for data analysis											
Attività 4.1	Valutazione della tipologia dei dati disponibili e definizione dei migliori algoritmi predittivi da sviluppare per la loro analisi		INFN	RI							
Attività 4.2	Produzione di set rappresentativi di dati complessi e loro successiva analisi sulla base degli algoritmi QML proposti										
Attività 4.3	Analisi dei dati ambientali sulla base degli algoritmi QML sviluppati										
OR5: IoT sensors network											
Attività 5.1	Modellazione hardware e progettazione elettronica dei sensori e dei sistemi di acquisizione		IBT	RI							
Attività 5.2	Progettazione dei protocolli di comunicazione IoT e del firmware di gestione										
OR6: Infrastruttura dati e Control Room											
Attività 6.1	Definizione Architettura Control Room		ARPA	SS							
Attività 6.2	Implementazione Virtual Machine - Virtual Network - Virtual Storage Control Room										
Attività 6.3	Integrazione Control Room - Ecosistema ICT ARPA Sicilia										
Attività 6.4	Implementazione Piattaforma di Monitoraggio Architettura Control Room										
Attività 6.5	Analisi Efficacia Architettura e Tuning Dimensionamento - SLA (Service Level Agreement)										
OR7: Sviluppo environmental intelligence platform											
Attività 7.1	Implementazione algoritmi di scientific advanced analytics e Artificial intelligence (servizi)		CORV	SS							
Attività 7.2	Implementazione Architettura Environmental intelligence platform										
Attività 7.3	Implementazione Interfaccia utente Environmental intelligence platform										
OR8: Data Hub											
Attività 8.1	Specifica requisiti ed architettura Data Hub		IBT	SS							
Attività 8.2	Sviluppo prototipo Data Hub										

INFN-LNS Contribution and QML

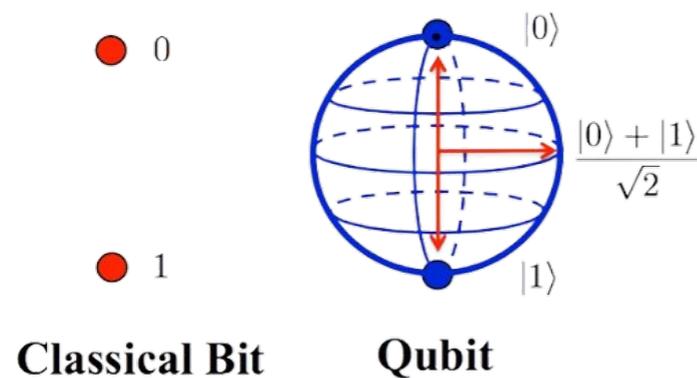
Beyond common logic : quantum computers

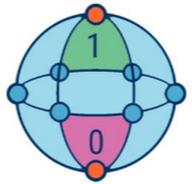
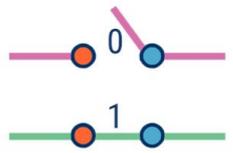
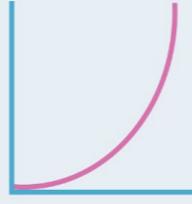
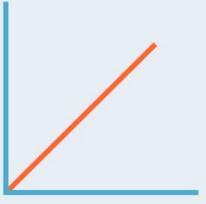
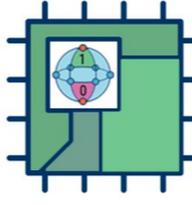
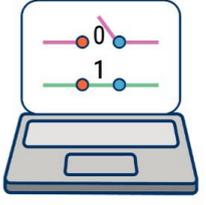
Bits

- The device computes by manipulating the bits through logical gates
- Classical computers has a memory made of of collections of 1 or 0 bits
- Gates, bits are implemented through the use of transistors

Qubits

- The device computes by manipulating qubits through the use of quantum gates
- A qubit can be 1, 0 and a superposition of those

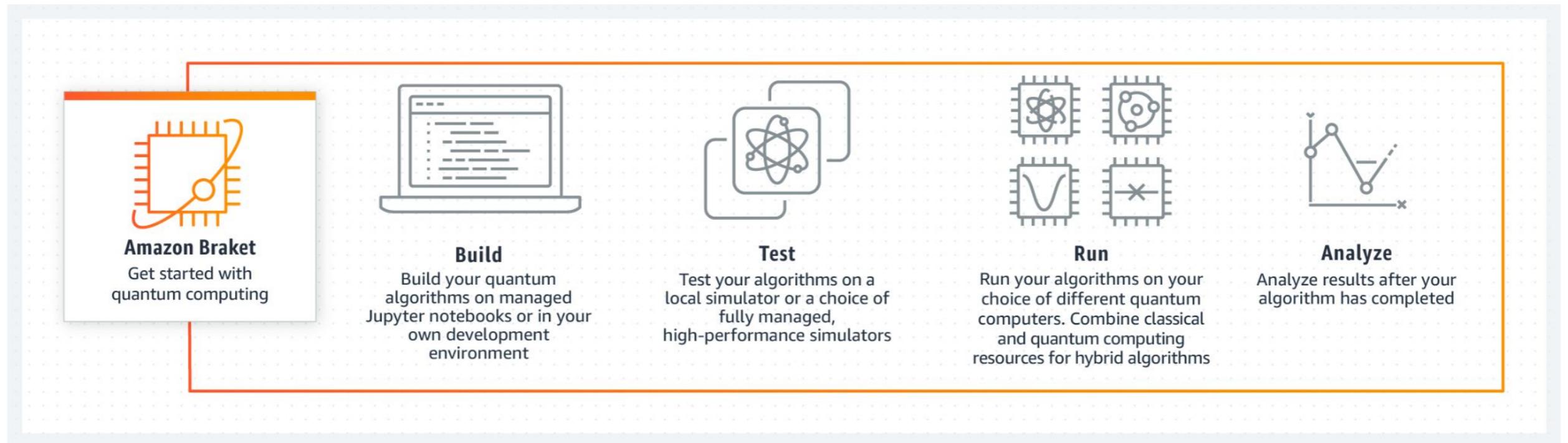


Quantum Computing	Vs.	Classical Computing
 <p>Calculates with qubits, which can represent 0 and 1 at the same time</p>		 <p>Calculates with transistors, which can represent either 0 or 1</p>
 <p>Power increases exponentially in proportion to the number of qubits</p>		 <p>Power increases in a 1:1 relationship with the number of transistors</p>
 <p>Quantum computers have high error rates and need to be kept ultracold</p>		 <p>Classical computers have low error rates and can operate at room temp</p>
 <p>Well suited for tasks like optimization problems, data analysis, and simulations</p>		 <p>Most everyday processing is best handled by classical computers</p>

INFN-LNS Contribution and QML

Amazon BraKet Partnership

Amazon BraKet è un servizio di calcolo quantistico completamente gestito progettato per aiutare a velocizzare la ricerca scientifica e lo sviluppo di software per il calcolo quantistico.



Press Release 2021

17 JUNE 2021

THE ITALIAN NATIONAL INSTITUTE FOR NUCLEAR PHYSICS AND AMAZON WEB SERVICES
WORK TOGETHER TO ACCELERATE QUANTUM COMPUTING RESEARCH

INFN-LNS Contribution and QML

Xanadu PennyLane

A cross-platform Python library for differentiable programming of quantum computers. Train a quantum computer the same way as a neural network.

Quantum Machine Learning

examples already implemented in PennyLane useful for our interests:

- Variational Classifier
- Quantum Generative Adversarial Networks
- Approximation of Classical ML Kernels on Quantum hardware



Scopo dell'OR4 è quello di **valutare in modo predittivo**, con un approccio ibrido (**unendo strumenti di simulazione analitici convenzionali, approcci di calcolo basati su tecniche Machine Learning (ML) e calcolo quantistico**), e utilizzando come punto di partenza i dati ambientali forniti da ARPA poi memorizzati, normalizzati e georeferenziati nella nuova piattaforma informatica, **il danno subito da una cellula a livello del DNA e riuscire a ricostruire la sua successiva evoluzione in un tessuto o una forma patologica**. Tutto ciò sarà fatto anche con l'obiettivo di poter prevedere nei prossimi anni l'insorgenza di malattie alla popolazione esposta al particolare agente esterno.

Oggi, pur disponendo di avanzati sistemi di calcolo, basati su algoritmi in grado di riprodurre in modo preciso il danno fisico e chimico prodotto a livello della cellula e delle sue strutture, **non disponiamo, a causa della estrema complessità dei sistemi in gioco, di strumenti capaci poi di ricostruire l'evoluzione della stessa cellula a seguito del danno**. Questo rappresenta oggi, un obiettivo limite nella previsione di un danno e il **colmare questa conoscenza rappresenterebbe un breakthrough** non solo nella strada per la comprensione dei fenomeni neoplastici ma, in generale, nelle previsioni della evoluzione di una singola cellula nei sistemi complessi che caratterizzano i sistemi biologici viventi. Il cercare di colmare parte di questi limiti con lo studio e implementazione di approccio di calcolo innovativi basati sul Monte Carlo, il Machine Learning e il calcolo quantistico, è l'obiettivo di questo OR.

Il nostro obiettivo principale è l'impostazione **CQ del quantum machine learning, che utilizza il calcolo quantistico per elaborare set di dati classici**. I dati possono essere costituiti da qualsiasi tipo di osservazione da un sistema tradizionale. Di conseguenza, l'obiettivo principale dell'approccio CQ è creare algoritmi quantistici per il data mining applicabili anche al caso di quantità di dati molto grossi.

Quando si addestra un modello di machine learning, **si è di solito interessati a trovare il modello "più piccolo" (meno complesso) che può effettivamente catturare i pattern contenuti nei dati**. Il motivo di ciò è una proprietà nota dei modelli statistici indicati come **bias variance tradeoff** (Hastie et al., 2009). Una motivazione di questo aspetto potrebbe essere la seguente: **i modelli di ML possono riconoscere schemi all'interno dei dati. Una volta che questi schemi sono stati riconosciuti, questi possono anche produrre previsioni in base ad essi, in base ai dati che vengono forniti**. Ricerche recenti (Arute et al., 2019) mostrano che **i computer quantistici possono campionare da distribuzioni di probabilità che sono esponenzialmente più complesse da campionare rispetto ai computer classici**.

Per sviluppare gli algoritmi di predizione del danno dai dati ambientali, **partiremo dal complesso danno subito da una cellula e dal suo DNA da parte di una radiazione esterna.** I dati complessi così ottenuti saranno poi analizzati con approcci ML e di quantum computing. La radiazione esterna sarà modellizzata attraverso il codice Monte Carlo **Geant4-DNA.**

Le interazioni di tipo fisico chimico e biologico della radiazione con la materia organica sono intrinsecamente stocastiche, quindi per la loro modellazione è richiesto un modello stocastico. Una conseguenza diretta del miglioramento delle performance e della disponibilità di hardware per computer è che si sta verificando una progressiva transizione dai modelli analitici a più realistici modelli stocastici (i.e. Monte Carlo) per la simulazione di tali processi. **Dal 2007 GEANT-4 DNA offre la possibilità di descrivere le interazioni delle particelle in acqua su scala nanometrica in maniera più efficiente dei modelli ab-initio. I modelli presenti nel pacchetto descrivono principalmente le interazioni elettromagnetiche evento-per-evento di particelle (elettroni, protoni e atomi di idrogeno neutri, particelle alfa compresi i loro stati di carica) con l'acqua (oltre alla ionizzazione per alcuni ioni – Li, Be, B, C, N, O, Si e Fe).** In breve, i processi elettronici includono ionizzazione, eccitazione elettronica, scattering elastico, eccitazione vibrazionale e attaccamento molecolare (Incerti et al., 2016).

DELIVERABLE D 4.1.1

Report con la tipologia di algoritmi QML più adatti da adoperare per l'analisi predittiva dei dati Attività 4.2 - Produzione di set rappresentativi di dati complessi e loro successiva analisi sulla base degli algoritmi QML proposti Come esempio di danno cellulare, poi estensibile a qualunque tipo di dato verrà utilizzato quello, diretto e indiretto (di tipo chimico) provocato alla doppia elica del DNA umano da una radiazione ionizzante ad alto LET (Linear Energy Transfer) come uno ione idrogeno. Tale danno verrà simulato usando il tool Monte Carlo Geant4-DNA e il danno alla molecola quantificato in termini di rotture chimiche. La simulazione di un tale effetto prevede il tracciamento di un numero enorme di specie chimiche, provenienti dalle reazioni di idrolisi dell'acqua e la simulazione sia delle loro interazioni con le molecole costituenti il DNA che tra loro stesse. La complessità di tali processi richiede enormi tempi di calcolo che ne limitano la reale applicabilità. Verranno quindi, in questo stadio, sviluppati specifici approcci ML capaci di riprodurre il danno cellulare in modo generico partendo da un addestramento basato su simulazioni Monte Carlo complete. Gli algoritmi ML verranno quindi risolti usando approcci di calcolo quantistico come descritto.

DELIVERABLE D 4.2.1

Report con i risultati ottenuti applicando approcci di quantum computing su analisi dati di tipo ML e applicati a set di dati complessi di danno al DNA ottenuti attraverso simulazioni Monte Carlo Attività 4.3 - Analisi dei dati ambientali sulla base degli algoritmi QML sviluppati Software analitici verranno adoperati per riprodurre la crescita di un tessuto cellulare partendo dalla singola cellula danneggiata. Tali software sono in grado di simulare gli schemi riproduttivi cellulari sulla base dei complessi sistemi di comunicazione cellulare che includono processi chimici e chimico-fisici. Anche in questo caso, l'enorme complessità alla base di tali processi, oggi di fatto limita la possibilità di simulare l'evoluzione di una cellula danneggiata in un eventuale tessuto patologico (tumorale o in generale, a fenotipo patologico). Nel progetto si intendono quindi applicare algoritmi QML anche nel caso dello sviluppo di una cellula in un tessuto più complesso con l'obiettivo di coprire il gap oggi esistente nella capacità di collegare il danno al micro/nano-metrico al DNA e il successivo sviluppo di una patologia macroscopica.

DELIVERABLE D 4.3.1

Report sui risultati finali ottenuti gli algoritmi QML sui dati ambientali finali

Project Resilio

INFN

PIANO FINANZIARIO PER LA REALIZZAZIONE DEL PROGRAMMA

FABBISOGNO	Anno (2023)	Anno (2024)	Anno (2025)	Anno (2026)
Spese previste ammissibili	€ 40.000,00	€ 64.000,00	€ 56.000,00	€ 0,00
IVA	€ 8.800,00	€ 14.080,00	€ 12.320,00	€ 0,00
Totale	€ 48.800,00	€ 78.080,00	€ 68.320,00	€ 0,00

INFN

Attrezzature e strumentazioni

Non sono previste spese riconducibili a questa tipologia.

Consulenze e prestazioni

Nell'ambito del progetto sarà necessario disporre di adeguato accesso a macchine quantistiche. Il budget previsto sarà principalmente dedicato ad acquistare questo tempo macchina. Il budget sarà anche adoperato per alcune license software. In particolare si prevede l'acquisto di Matlab nella versione completa. Matlab è necessario per fare parte del benchmark dei risultati che si otterranno.

Materiali

Verranno acquistati consumabili per il laboratorio di radiobiologia che serviranno per i test di crescita dei tessuti che prenderemo come "modello" preliminare dei dati da analizzare e per il supporto all'uso dell'irradiare a Raggi X.