ELIXIR: Infrastruttura di ricerca europea per le Scienze della Vita

Graziano Pesole, ELIXIR-IIB Head of Node

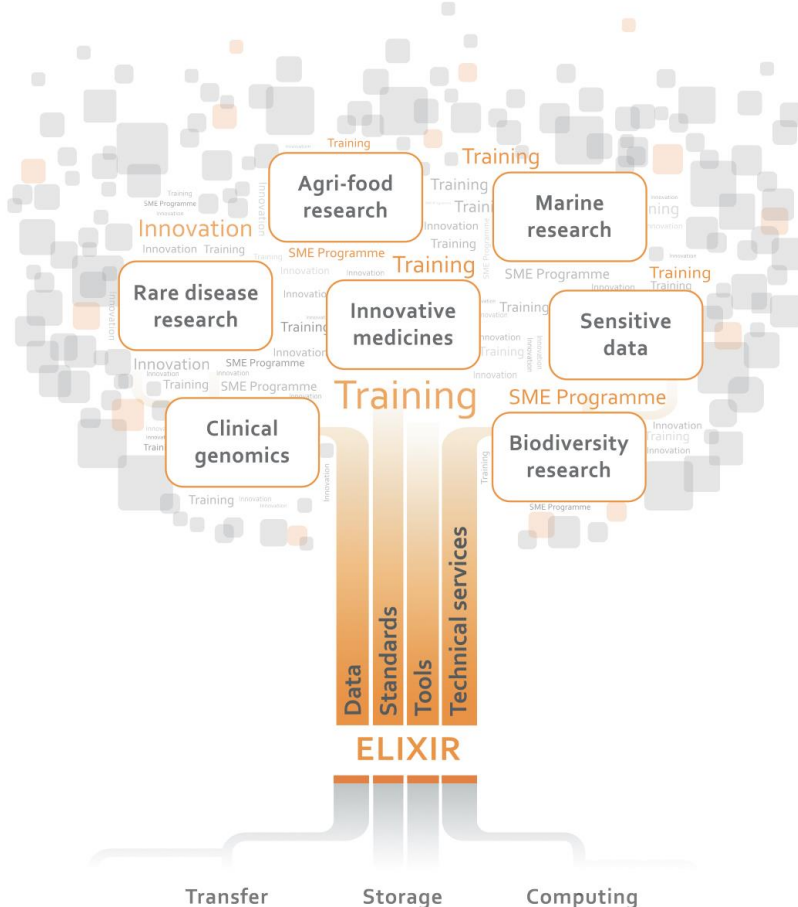CNR-IBIOM and University of Bari

Roadmap of the High Performance Computing in Italy and the settlement of a Scientific Advisory Board in CINECA

CINECA, Bologna 23 March 2018

ELIXIR: A pan-european distributed Infrastructure for Bioinformatics

ELIXIR (www.elixir-europe.org) is a European Research Infrastructure for biological data whose primary objective is to support research in the field of "life sciences" and their translational activities to medicine, environment, biotechnological industries and society.
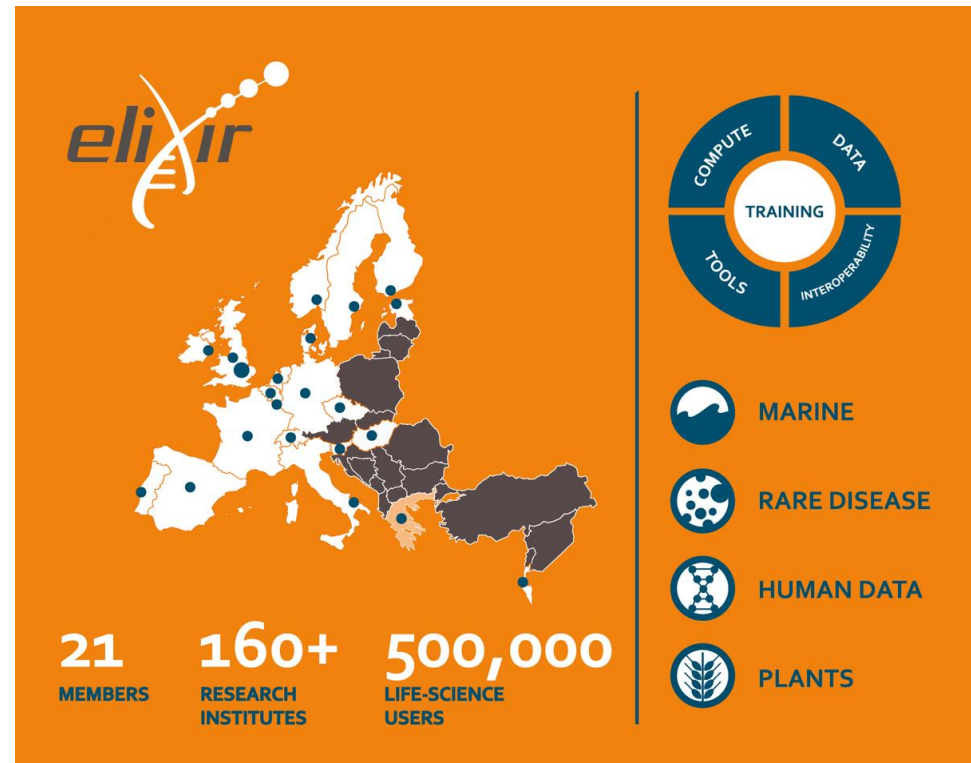

ELIXIR is an intergovernmental organisation, whose legal framework is an **EMBL special project**, that brings together life science resources from across Europe. These resources include databases, bioinformatic software tools, training materials, cloud storage and supercomputers.

Since **January 2014** ELIXIR is a permanent legal entity and is now considered a "**Landmark Research Infrastructure**".

ELIXIR: A pan-european distributed Infrastructure for Bioinformatics

ELIXIR is structured as a central hub, located in the Wellcome Genome Campus (Hinxton, UK) and 20 national nodes including over 160 Research Organizations.

ELIXIR Overview 4

ELIXIR: A pan-european distributed Infrastructure for Bioinformatics

ELIXIR Platforms

COMPUTE - Develop ways that researchers across Europe can access, store, transfer and analyse large amounts of life science data.

DATA - Identify key data resources across Europe and support the linkages between data and literature e.g. by making it easier to move from a scientific paper to the dataset that the paper was based on.

TOOLS - Provide ways for researchers to find the best software to analyse their data.

INTEROPERABILITY - Establish Europe-wide standards that can be used to describe life science data. This makes different data sets easier to compare and analyse.

TRAINING - Help scientists and developers find the training they need, and also provide it.
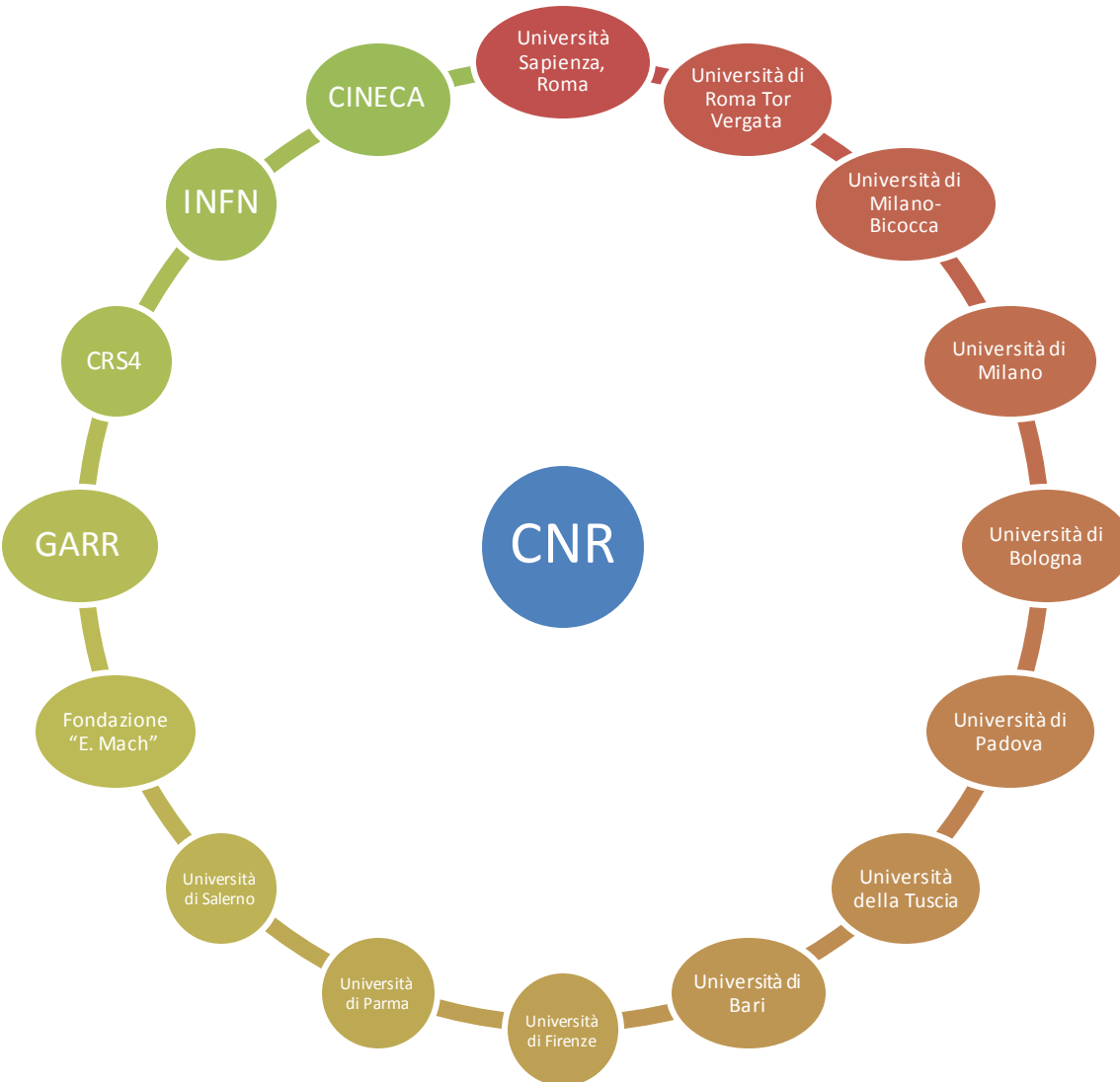
ELIXIR Use Cases

HUMAN DATA - Develop ways that sensitive data from biomedical research can be discovered and accessed by scientists.

RARE DISEASES - Work with researchers, clinicians and industry to promote the development of new therapies for rare diseases.

MARINE METAGENOMICS - Develop standards, databases, tools and training to help those studying the genetic material sampled from the sea.

PLANT SCIENCES - Develop standards and tools to help those studying the genetics of crops and trees.

The ELIXIR Italian Node

The Italian node is configured as a **Joint Research Unit** (JRU) named **ELIXIR-IIB** which is in charge of coordinating the delivery of existing bioinformatics services at the national level, also pursuing their integration in the ELIXIR infrastructure.

ELIXIR-IIB is led by National Research Council (CNR) of Italy and comprises other 16 partners including several universities as well as leading high-performance computing partners such as CINECA, CRS4, GARR and INFN.

The ELIXIR Italian Node

Sistemi di supercalcolo

- CINECA/Marconi: Convenzione ELIXIR-IIB HPC@Cineca

- INFN: Accordo nell'ambito della JRU ELIXIR-IIB

Applicativi e Modelli di simulazione

- Vari applicativi bioinformatici, ad es: allineamento sequenze NGS, assemblaggio genomi, variant calling, predizione strutture RNA e proteine…

Tipologie di servizi innovativi

- CLOUD HPC con particolare riferimento alla fornitura di servizi bioinformatici su cloud a realtà di ricerca o cliniche. In particolare, per queste ultime è importante affrontare al più presto il tema dell'accesso controllato ai dati (GDPR).

Modelli di utilizzo

- Entrambi i modelli di utilizzo: i) interattivo (es. Pipeline di riduzione di grosse mole di dati con interazione non frammentata nel tempo); ii) batch (workflow pianificabile in specifici step senza impatto nel time-to-science per lo specifico caso) sono importanti e si prestano ad applicazioni bioinformatiche diverse.

Modelli futuri

- Gli sviluppi futuri dovranno misurarsi con le necessità che possono già essere prefigurate in termini di capacità di storage (100 Pb nei prossimi 5 anni) e di calcolo. I servizi basati su Cloud e HPC cloud dovranno rispondere a requisiti fondamentali quali l'elasticità, la sicurezza e l'accesso controllato ai dati.

Competenze

- Personale tecnologico/ricercatore con competenze specifiche di dominio (Bioinformatica)

Modelli di accesso

- Modelli a sportello cloud like che offrano piattaforme e ambienti ready-to-use