

ELIXIR: European Life Science Infrastructure for Biological Information



Graziano PESOLE

University of Bari & IBBE-CNR, Bari, Italy
Elixir JRU Manager

A banner for the "SUD LABORATORIO DIGITALE" workshop. The top left features the "GARR X PROGRESS" logo. To its right is a navigation menu with links: "Workshop GXP", "Sede e Trasporti", "Programma", "Registrazione", "Streaming & WiFi", and "GARR-X Progress", followed by a search icon. Below the menu, a green bar indicates "Workshop • Palermo 31MAR-02APR". The main part of the banner has a light gray background with a collage of colorful icons representing various digital and scientific concepts. On the right, the text "Palermo 31 marzo- 2 aprile" is displayed in a large, white, sans-serif font, followed by "SUD" in green and "LABORATORIO DIGITALE" in large, bold, blue and orange letters.

ELIXIR was identified in 2006 as a priority by the European Strategy Forum on Research Infrastructures (ESFRI), and under the Seventh Framework Programme, the European Commission has funded a project to promote the preparatory stages of such an infrastructure, completed at the end of 2011, which was also participated by the National Research Council through the former Department of Life Sciences.

Juridical model of governance:

EMBL Special project

Construction process:

Memorandum of Understanding (MoU)

- 17 countries (Italy, July 2012)

On **18 December 2013**, ELIXIR became a permanent legal entity following the ratification of the ELIXIR Consortium Agreement by EMBL and the first five countries. The countries that have signed the ECA (UK, Sweden, Switzerland, Czech republic, Estonia, Norway, Netherlands, Denmark) are full members of the ELIXIR Board.



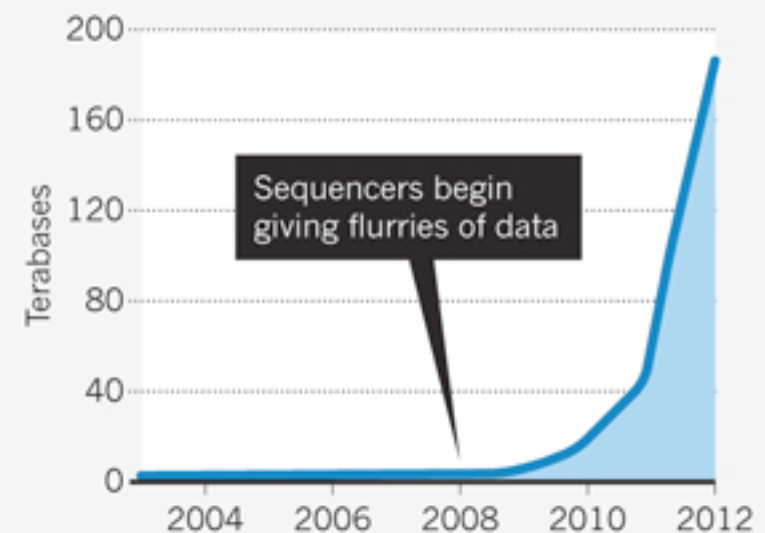
The mission of ELIXIR is building a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society.

The challenge that biomolecular data resources face today is two-fold. Firstly, new technologies such as next-generation DNA sequencing are generating massive amounts of data. It has been estimated that between today and 2020 these new technologies will produce data at up to one million times the current rate. Secondly, there is an emerging and pressing need to provide infrastructure that will meaningfully integrate new types of data to be collected in the future.

The collection, curation, storage, archiving, integration and deployment of biomolecular data is an immense challenge that cannot be handled by a single organisation or by one country alone, but requires international coordination. Over recent years European countries have invested heavily in research that produces these data. It is now recognised that there is an urgent need for a pan-European infrastructure that will facilitate the process of extraction of optimum value from current and planned investments in this area.

DATA EXPLOSION

The amount of genetic sequencing data stored at the European Bioinformatics Institute takes less than a year to double in size.



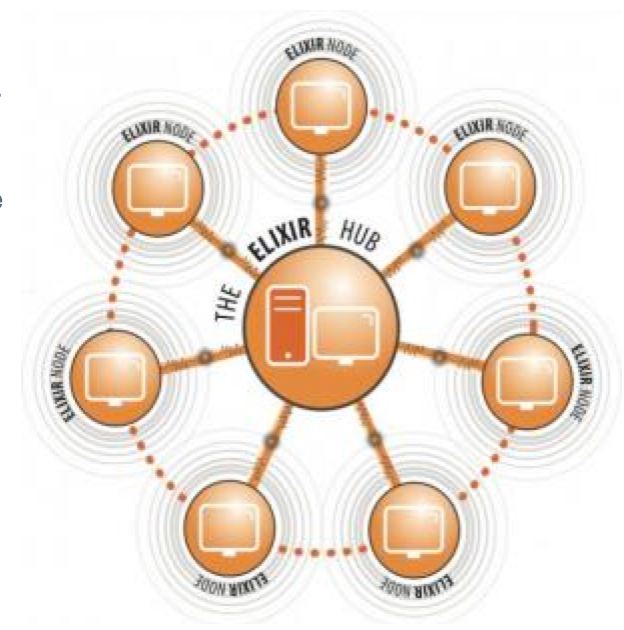
ELIXIR: European Life Science Infrastructure for Biological Information



ELIXIR will be built as a distributed infrastructure on different nodes hosted by centers of excellence located throughout Europe (the "ELIXIR nodes"); these nodes will be connected to a central hub ('ELIXIR Hub') which is located at the European Bioinformatics Institute EMBL (EMBL-EBI) in Hinxton, UK.

The ELIXIR Hub will host the Management Executive Committee and the Secretariat of ELIXIR. It will coordinate the services of the "data center" run by EMBL-EBI and the coordination of services managed by ELIXIR nodes. It will also provide the basic data (the data that is essential for almost all areas of research in the life sciences) to users.

Peripheral ELIXIR nodes will provide a range of specialized services and complementary to those provided by the central node. The integration of ELIXIR Nodes will provide essential components distributed by the ELIXIR infrastructure at European level.



Next-Generation Sequencing



A large number of platforms using different strategies and chemistries, and with a differently high throughput are coming in the market.



Roche / 454 Genome
Sequencer FLX titanium
(800 bp, 800 Mb / run)



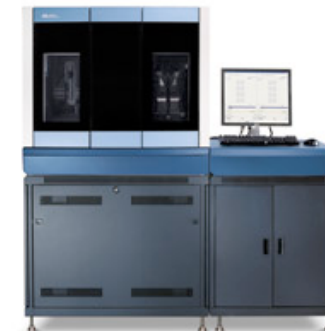
Ion Proton



Illumina / Solexa
Genetic Analyzer HiSeq 2000
(150x2 bp, 600 Gb / run)

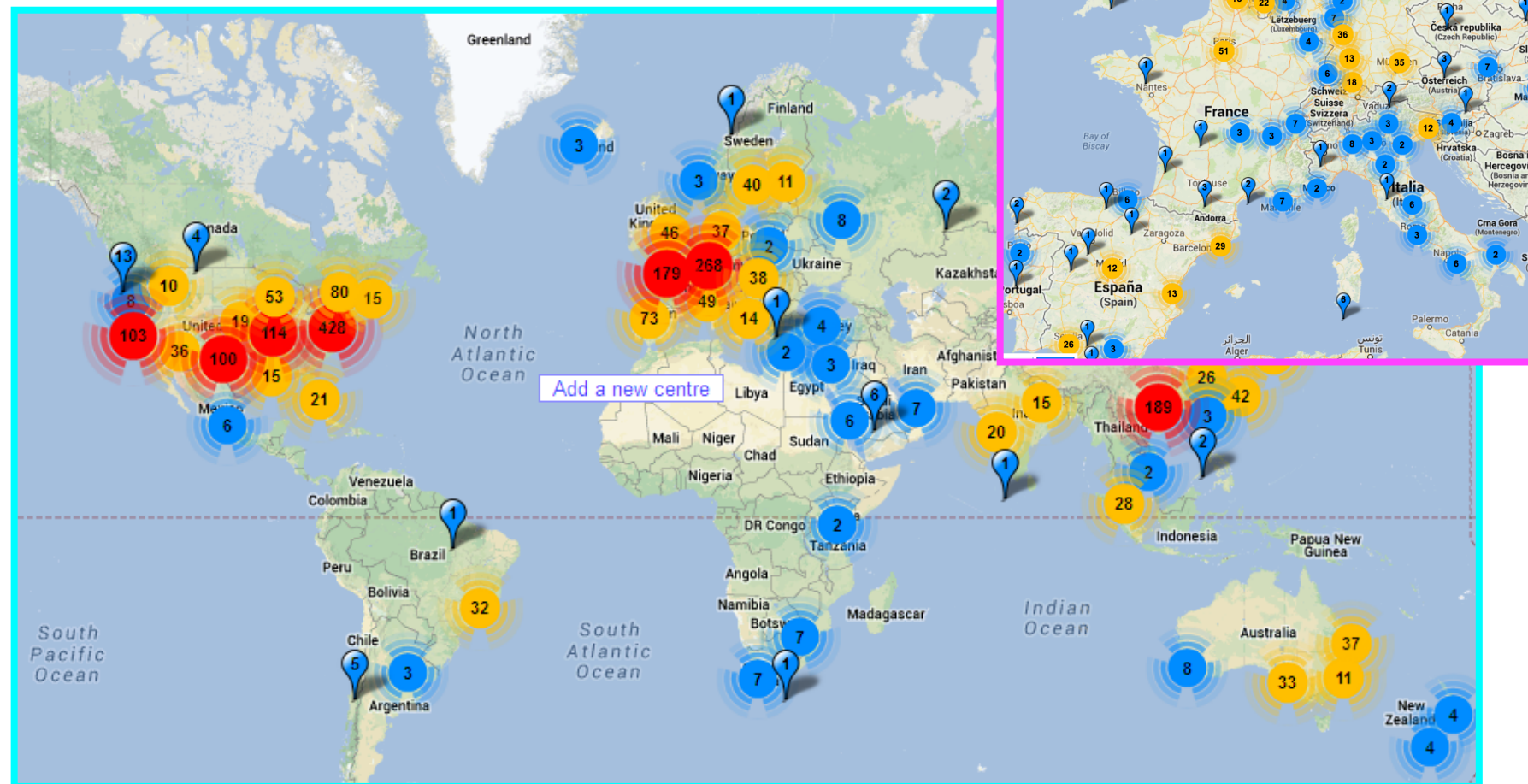


PacBio



Applied Biosystems
SOLiD 4 System™
(100x2 bp, 400 Gb / run)

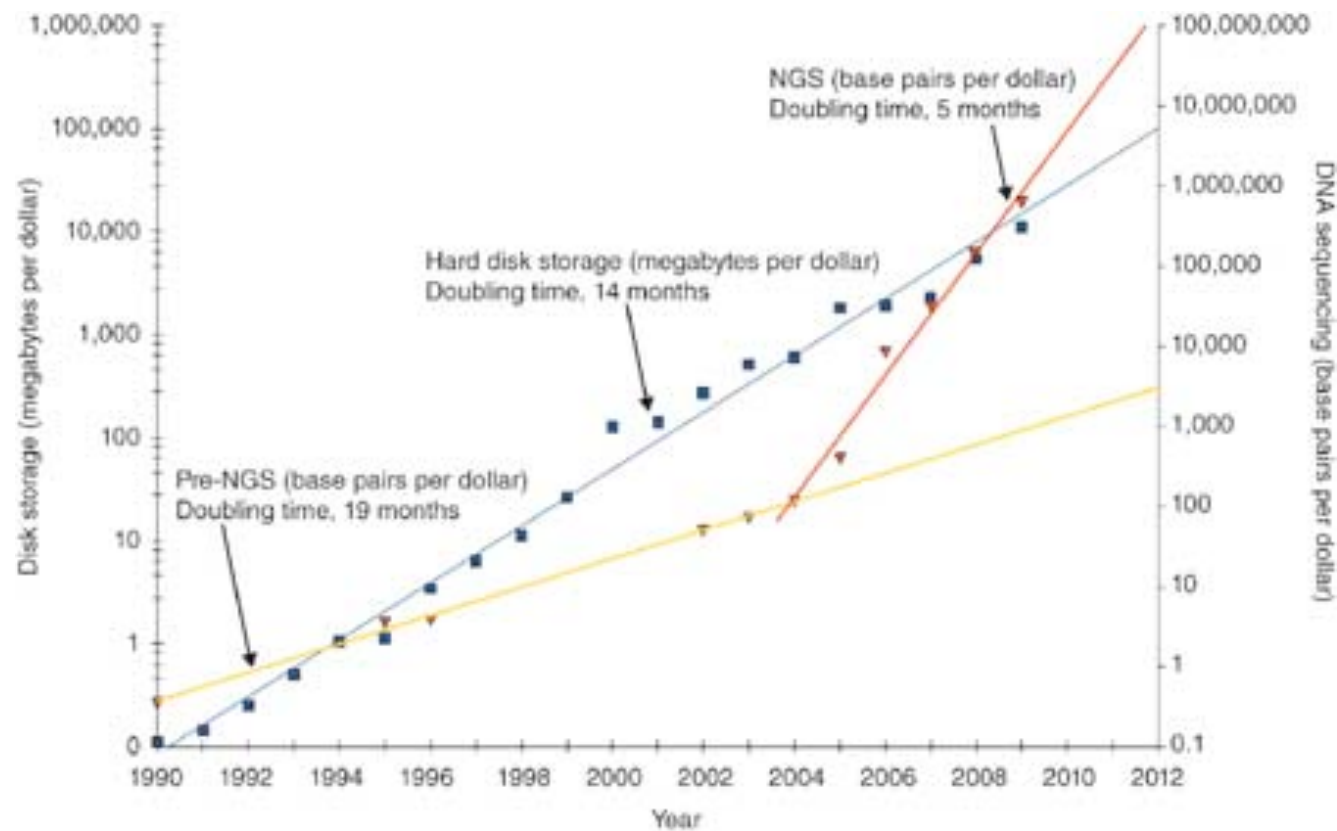
Worldwide distribution of NGS platforms



>2558 total machines (>1000 US, >300 China, .., >50 Italy)

Source: omicsmap.com

Next-Generation Sequencing: data growth & costs



The ELIXIR italian node

In order to build the Italian node of ELIXIR infrastructure a Joint Research Unit has been established, coordinated by the National Research Council, which defines its scope and methods of governance, with the participation of research institutions, academic and technological as listed below:

- 1) Consiglio Nazionale delle Ricerche (G. Pesole, A. Facchiano, L. Milanesi)
 - 2) Università di Roma "Sapienza" (A. Tramontano)
 - 3) Università di Roma "Tor Vergata" (G. Cesareni)
 - 4) Università di Milano-Bicocca (G. Mauri)
 - 5) Università di Bologna (R. Casadio)
 - 6) Università di Padova (G. Valle)
 - 7) Università di Milano (G. Pavesi)
 - 8) Università della Tuscia (A. Valentini)
 - 9) CRS4 (G. Zanetti)
 - 10) GARR (E. Valente)
 - 11) CINECA (S. Bassini)
 - 12) INFN (G. Maggi)
- Universit /EPR
- Technological Provider



The ELIXIR italian node

The initial nucleus of the institutions that formed the JRU includes all those who had submitted organic proposals for the involvement in ELIXIR in responding to a "call for expression of interest" in the summer of 2010. These proposals had been positively evaluated by the "Steering Committee" of ELIXIR.

Elixir-Ita Steering Committee

JRU Manager (Head of Node):	Graziano Pesole
Italian delegate of Elixir Interim Board:	Anna Tramontano
Elected member (Chief of Admin Affairs):	Giancarlo Mauri
Elected Member:	Giorgio Valle
Elected member	Sanzio Bassini



Italy Node sets up as JRU

In August 2013, after the official signature of the MoU, Italy formally established its Joint Research Unit (JRU), which will act as the legal entity for the Italian ELIXIR Node. Led by Professor Graziano Pesole, Director of the Institute of Biomembranes and Bioenergetics in Bari, the ELIXIR Node is hosted by CNR, which is the National Research Council. In addition to CNR, the Italy ELIXIR Node also brings together 12 partners including several universities as well as leading high-performance computing partners such as CINECA, CRS4, GARR and INFN. The ELIXIR Italy Node has also established a robust procedure, based on an open call and a peer review system, to allow additional participants to join.

Italian proposals to contribute to ELIXIR's construction (2010)

Proposer	Description
Gianni Cesareni, Univ. Tor Vergata, Rome	Protein – protein interaction resources (MINT)
Anna Tramontano, Univ. Sapienza, Rome	Training in Bioinformatics
Giorgio Valle, University of Padua	Plant Genomics resources
Gianluigi Zanetti, CRS4, Cagliari	Human Genomics resources
Alessio Valentini, Univ. Tuscia, Viterbo	Farm Animal Genomics resources
Graziano Pesole, IBBE-CNR, Bari	Transcriptome resources and tools

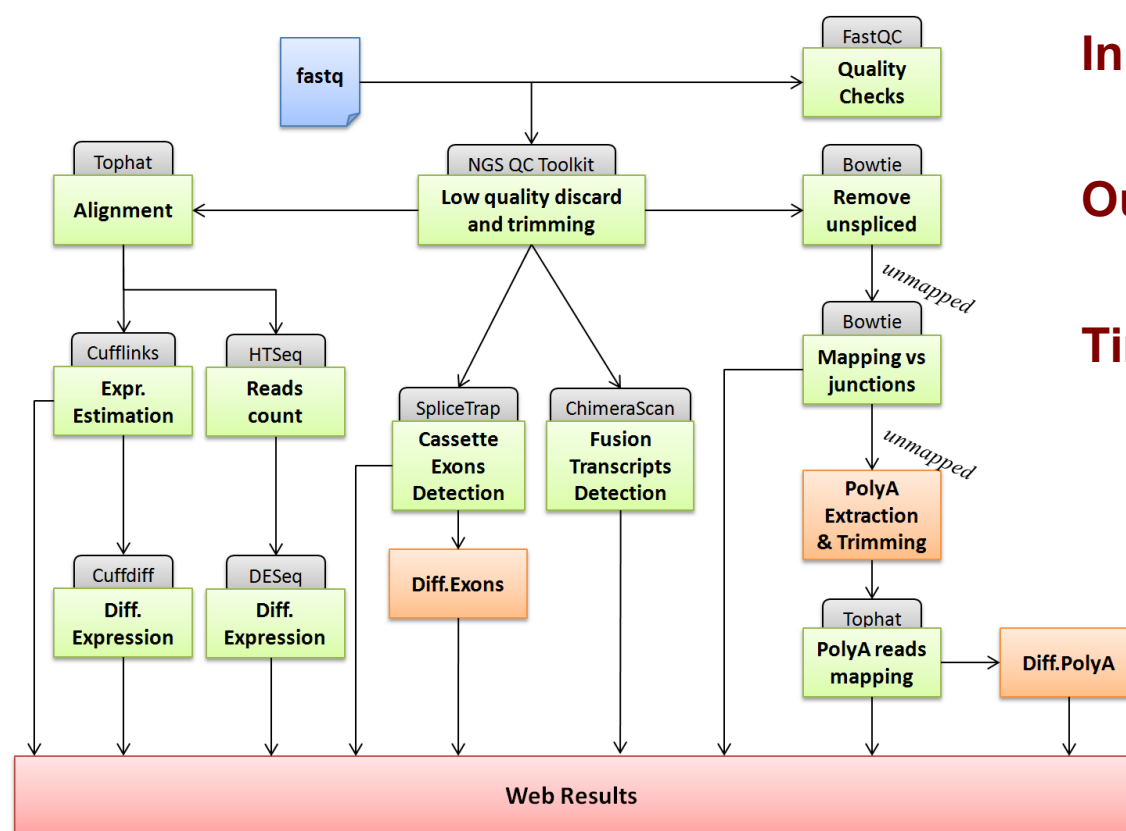
Italian Node Resources

Description	Web address
ASPicDB (alternative splicing patterns in the human and other genomes)	http://srv00.ibbe.cnr.it/ASPicDB/
SpliceAid-F (splicing regulatory factors and their binding sites)	http://srv00.ibbe.cnr.it/SpliceAidF/
UTRdb/UTRsite (eukaryotic mRNA untranslated regions and their regulatory elements)	http://www.ba.itb.cnr.it/UTR/
Mitozoa (database of comparative mitogenomics in metazoans)	http://srv00.ibbe.cnr.it/mitozoa/
ITSoneDB (database of fungal ribosomal RNA Internal Transcribed Spacer sequences)	http://itsonedb.ba.itb.cnr.it:8080/ITS1/
DIGIT (database of immunoglobulin variable sequences)	http://biocomputing.it/digit/
MAISTAS (modeling and assessment of splicing isoforms)	http://maistas.bioinformatica.crs4.it/
MobiDB (database of protein disorder and mobile regions)	http://mobidb.bio.unipd.it/
PIGS (prediction of Immunoglobulin structures)	http://circe.med.uniroma1.it/pigs/
The grape genome browser	http://genomes.cribi.unipd.it/

RNA-Seq Analysis Workflow



All NGS data analysis require adequate IT infrastructures and are computationally very intensive. An example



Input: FASTQ (50 – 250 Gb)

Output: various (100 – 750 Gb)

Time: 70 – 100 h

CINECA infrastructure:

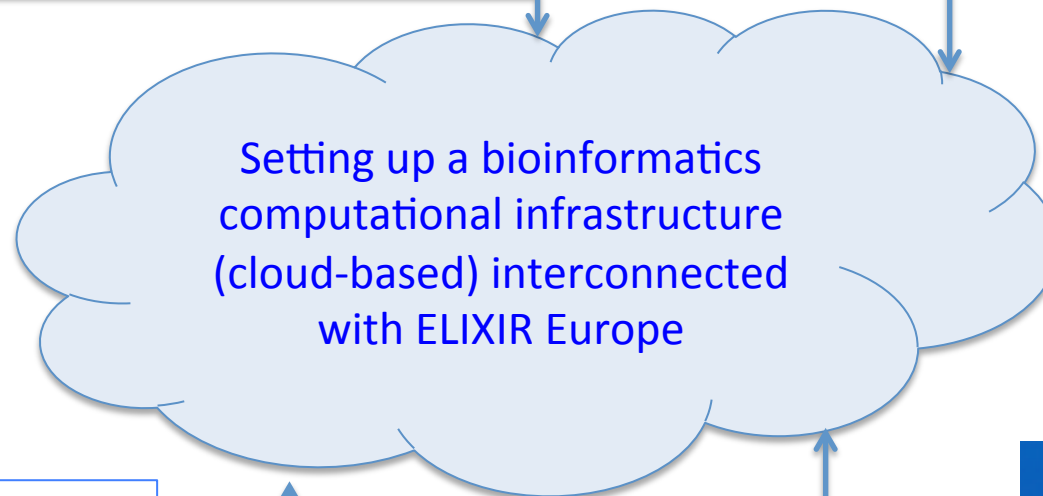
Cluster with 274 IBM iDataPlex M3 nodes (2 six-cores Intel Westmere at 2.40 GHz and 48GB of RAM)
Server HP DL980 8 Quad-Core Intel Xeon at 1.87 GHz with 512 GB of RAM

The italian node of Elixir: future plans

After the start-up phase, relative to the first year of operation, we plan to enhance the activities of ELIXIR-ITA, as shown below:

- Strengthening of the Organizing Secretary and activation of the management bodies of the JRU.
- Registry of bioinformatics resources available nationwide (databases, tools for data analysis, computational resources) that meet the requirements by Elixir.
- Activation of an organic link with the Elixir hub, at the European Bioinformatics Institute (EBI), through the figure of the NTO, with the task of defining standards of operability and integration of resources.
- Implementation of training activities in collaboration with EBI and other Elixir national nodes.
- Coordination activities with other national initiatives of interest for the bioinformatics infrastructure (e.g. flagship national projects., etc.).
- Initiation of actions for the creation of a distributed National Institute of Bioinformatics.

ELIXIR-ITA: Technological Partners



Call for the ELIXIR Italian Technical Coordinator

The main purpose of this role is to support the technical implementation of ELIXIR services in collaboration with other ELIXIR nodes to provide a sustainable European infrastructure for biological information. Thus the ELIXIR Italian Technical Coordinator will work closely with the Italian Head-of-Node, the technical coordinators from other ELIXIR Nodes and the Italian collaborators of ELIXIR Italy.

Key responsibilities:

- Lead the technical evaluation, development, monitoring, maintenance and sustainability of ELIXIR services.
- Advise the Italian Head-of-Node in establishing governance processes of direction and control to ensure that objectives are achieved.
- Report to the Italian Head-of-Node and the ELIXIR Hub Technical Coordinator
- Proactively work with ELIXIR nodes to identify, evaluate and implement appropriate technology platforms.
- Identify opportunities and risks for delivering ELIXIR services.
- Ensure that technology standards and best practices are maintained nationally and across the organisation.
- Identify and support prioritization of ELIXIR pilot actions and technical parts of ELIXIR grant applications.

Criteria to be included in the Elixir Italian Service Registry of Bioinformatic Resources

- Existing and running resource, and with a significant user base;
- Resource publicly available without restrictions;
- Ensuring sustainability for a sufficient period of time;
- Availability of appropriate documentation, according to a common standard agreed in the international framework.

ELIXIR: European Life Science Infrastructure for Biological Information

