

ELIXIR: a distributed infrastructure for European biological data

Lindsey C. Crosswell and Janet M. Thornton

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

The ELIXIR project

The most successful collaborative projects in the scientific panoply are those that provide solutions to problems shared by many worldwide. The ELIXIR project is one such example, its aim being to ensure that European countries are able to respond to the set of grand challenges that they all face. ELIXIR has as its stated mission 'the construction and operation of a sustainable infrastructure for biological information in Europe to support life science research and its translation to medicine and the environment, the bio-industries and society'.

To appreciate the enormity of this undertaking, it is helpful to take a step back to look at the problems that ELIXIR is seeking to address. Collectively, the European member states share the same complex range of challenges in the future; namely, a growing and ageing population requiring the development of new drugs to treat the conditions that these older citizens face. Alongside these demographic issues is the overriding necessity to secure food and water supply in an era of climatic change and environmental degradation [1].

Biological information is central to life sciences and biomedical research and therefore a robust and sustainable infrastructure to enable the curation, archiving and accessibility of core data is essential to support excellence in research and innovation in academic and industrial settings. Comparatively recent developments in high-throughput sequencing techniques have created a deluge of data that no single country alone can manage efficiently (Figure 1). We are entering an era in which an entire human genome can be sequenced within days and, in some cases within hours, at a cost of a few hundred Euros, a task that when begun nearly 20 years ago cost hundreds of millions of dollars, involved an international collaboration of scientists and took some 13 years to complete [2].

A united solution to the data deluge in Europe

The immense undertaking of gathering, curating, storing, archiving, integrating and deploying biomolecular data from disparate sources requires international coordination. ELIXIR, as one of 13 biomedical science research infrastructures that form part of the ESFRI (European Strategy Forum on Research Infrastructures) process, proposes a shared Europe-wide solution to safeguard the very significant investment that member states have already made and continue to make in generating life

sciences data, enabling them to extract maximum value from open access to shared core data. Extensive consultation with data users began in 2009 during the 4-year preparatory phase for ELIXIR funded by the EU Seventh Framework Programme (FP7). Information gathered on data access and user requirements was central in designing ELIXIR as a distributed infrastructure with a central hub located at the European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) at Hinxton, near Cambridge, UK. The hub will be connected to ELIXIR nodes to be hosted at centres of excellence in universities and institutions across Europe (Figure 2).

The next steps for ELIXIR

The announcement by the UK Large Facilities Capital Fund on 5 December 2011 that it has granted an award of £75 million for the construction of the hub building and the extension of the data centre signals the transition of ELIXIR from a virtual concept into a physical reality. To date, ten countries have signed the ELIXIR Memorandum of Understanding and of these, Denmark, Sweden, Finland, Spain and Norway have collectively pledged €22 million of investment towards construction of ELIXIR. Ministerial and scientific representatives from each of the signatory countries met as an interim board for the first time in late 2011 to agree the governance framework for the project.

The preparatory phase of the project pales into insignificance when set against the enormity of the task that lies ahead to construct ELIXIR and progress it to the point of full operability by 2016. Currently, the European bioinformatics landscape is largely fragmented across borders and institutions and to be technically effective and economically efficient, ELIXIR will incorporate nodes with unique functions or combinations of functions and competencies, including, for example, data provision, e-infrastructure provision or specialism in the fields of standards, tools and training. Ultimately, the success of ELIXIR will be measured by how indispensable it becomes as an application for scientists working in academia and industry.

The benefit of ELIXIR to the innovation economy in Europe

ELIXIR will build on current strengths across European nations, removing duplication, spreading best practice and, most importantly, enabling all data to be used to increase our knowledge of the molecular processes of life and disease. The vision for ELIXIR is for it to be permanently and openly available, robust in the face of

Corresponding author: Crosswell, L.C. (crosswel@ebi.ac.uk).

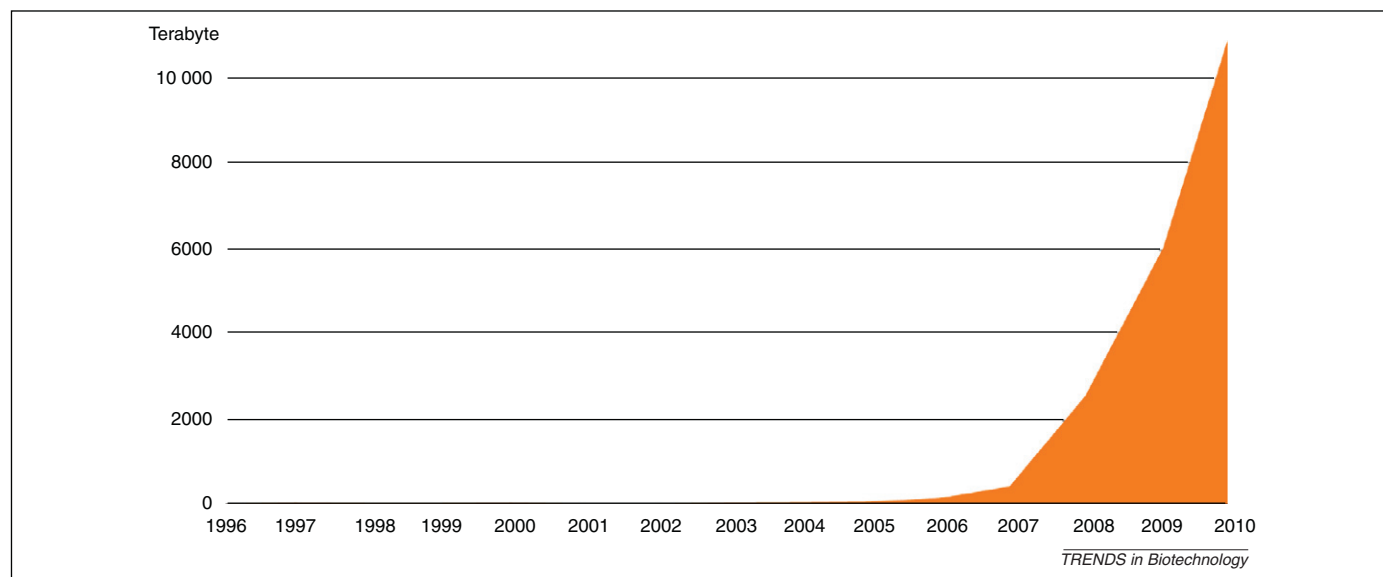


Figure 1. The scale of data growth. The chart shows the trend in storage capacity needed to store biological data at EMBL-EBI (a terabyte is a million million bytes).

external threats, dynamic, scalable and responsive to the rapidly increasing flow of data. However, it is not the technical design of ELIXIR alone that has persuaded European governments and their scientific funding bodies to make substantial investments to support the operating costs of the technical hub, and the national costs of the interconnected nodes, in a challenging prevailing economic climate. Instead, what European governments recognise as the true benefit of ELIXIR is the massive potential it has to unlock scientific innovation. Increasingly, as European countries abandon their traditional manufacturing heritage, they are seeking to replace this with pioneering research and innovation in biotechnology, molecular biology, genomics and other life sciences. It is the translation of this cutting-edge research into new discoveries and innovative developments in pharmaceutical, environmental and agribusiness companies, bringing the twin rewards of societal benefit and economic prosperity, that has led to the commitment of the member states to the delivery of ELIXIR as an irreplaceable resource for all.

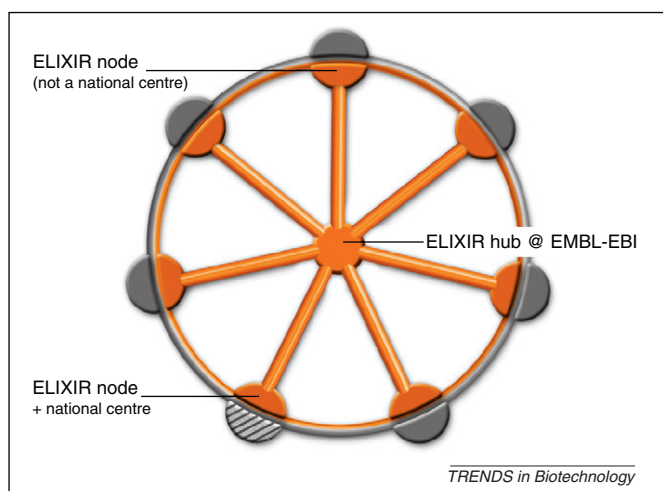


Figure 2. The ELIXIR hub and nodes structure. The figure shows how ELIXIR will be organised to complement the capabilities of its participating organisations. Striped shading illustrates an example in which an ELIXIR node is also a recognised national centre for excellence.

ELIXIR and personal genomics data

Governments are also planning ahead for the management of patient data as we enter the era of the personal genome. It is not unrealistic to consider that within this decade national health services in Europe will routinely sequence individual genomes, enabling treatment protocols to be devised with specific reference to the patient's genetic blueprint. Certainly the UK is on the cusp of an ethical shift whereby the Government intends to change the law in regard to the use of patient data in the development of clinical trials to enable them to be more widely used for vital research, while still protecting patient anonymity [3]. Other countries are also facing this challenge and are piloting various national projects for the management of patient data. ELIXIR will play a central role in the management of these data and potentially could help to contribute to such pilot efforts, becoming an exemplar as it optimises the use of medical and patient data for pharmaceutical innovation.

A European solution

The ELIXIR project is exciting precisely because it is not prescriptive about the future, it will create a mechanism for sharing of data on a common platform across national boundaries, and it will be adaptable to the era of substantial change in data volume and type that lies ahead. Without a doubt the value of ELIXIR to Europe will be vastly greater than the sum of its parts. To learn more about ELIXIR visit www.elixir-europe.org.

References

- 1 Nellesmann, C. *et al.* (2009) The Environmental Food Crisis, *The Environment's Role in Averting Future Food Crises*, UN Environmental Programme
- 2 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
- 3 Bloor, C. *et al.* (2011) *Strategy for UK Life Sciences*, Department for Business Innovation & Skills, Office for Life Sciences