



## 6 ELIXIR Domain Specific Services

**Work stream leads:** Alfonso Valencia (ES), Inge Jonassen (NO), Jose Leal (PT)

**Work stream members:** Nils-Peder Willassen (NO), Finn Drablos (NO), Mark Viant (UK), Ferran Sanz (ES), Morten Rye (NO), Luciano Milanesi (IT), Tiziana Castrignanò (IT), Paul Kersey (EMBL-EBI), Henning Hermjakob (EMBL-EBI)

### Remit

ELIXIR's preparatory phase analysed in detail the situation of Bioinformatics in various areas, including meetings with experts in academy and industry, in the various application areas, to determine their needs. Specific examples were presented and analyzed. The preparatory phase was also instrumental for the discussion of the infrastructures required in the main areas of bioinformatics and genomics.

The information of the preparatory phase has been instrumental for organization of the national nodes and shaping their proposal, and it is important for the organization of the ELIXIR domain specific services.

This documents is based on the results of the preparatory phase. Additionally, it includes aspects introduced by the node proposal and a general update of technical aspects that represent now pressing needs. Finally, this document outlines the principles for the collaboration of ELIXIR with domain specific communities in the framework of the h2020 program.

### Application domains

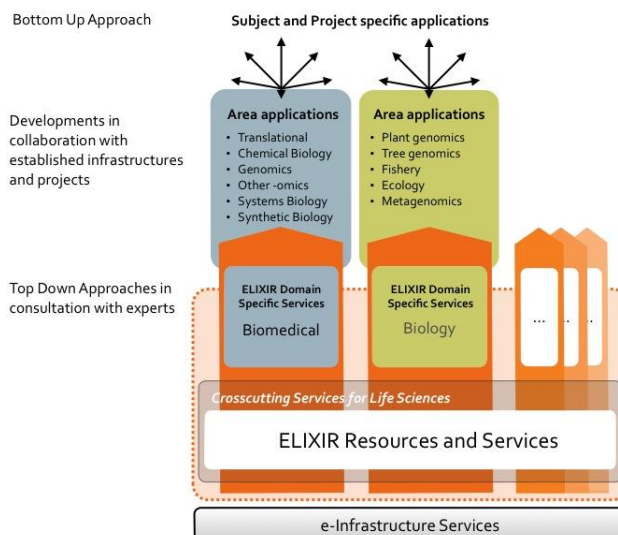
For the purpose of this document we divide the ELIXIR activities in two general themes: **Biomedicine** (including medical **biobanks**), and **Biology** (non-medical, including **agriculture, marine biology**, and industrial **biotechnology**). Both themes will have both commercial and non-commercial relevance. ELIXIR sees these areas from the perspective of genomics and other high throughput "omics" data sources specific of each domain.

The two areas can be subdivided in more specific domains. A non-exhaustive list of them includes:

*Biomedicine:* Translational bioinformatics and medical informatics, chemical biology, genomics and other "omics", systems and synthetic biology.

*Biology:* Plant genomics, tree genomics, fish genomics, marine biodiversity and ecology, metagenomics, bioprospecting, environmental genomics, (eco)toxicology and regulatory science and others.

In each one of those domains it is possible to identify a large number of initiatives, projects and alliances, including a number cases that will cut across domain boundaries.



## Actors and communities

- *ESFRI infrastructures.* A large number of the areas identified above as potential ELIXIR partners require their own infrastructures, and they are already organizing them in the ESFRI framework. The interaction with other ESFRIs, in particular with the BMS infrastructures, is an essential pillar of the ELIXIR construction.

The BioMedBridges project is a good example of how ELIXIR can work with BMS infrastructures. In this case the BMSs have obvious needs of appropriate data handling and analysis systems tailored to their scientific needs. ELIXIR can provide access to core methods and resources and gain from integration and common developments, always recognizing domain specific expertise and modes of working.

- *Innovative Medicines Initiative.* IMI represents a very large public/private partnership in the core of pharma industry. In this environment it is obvious that the domain specific projects will benefit from interrelation and common structural support. Indeed IMI already dedicates substantial resources to projects on harmonization and interoperability. The OpenPhacts project is a good example of this. Still all depends on access to core biomedical/molecular resources and their adaptation of specific needs of the IMI projects (we can mention eTOX toxicogenomics as one of those efforts).

- *Large scale genome projects,* including consortiums such as ENCODE, ICGC, iHEC and others, as they develop their own dynamics and represent an impressive concentration of scientific and technical expertise (in many cases including ELIXIR partners). These projects offer an excellent connection to international scenarios and developments. As in the other cases above, these projects largely rely on access to core resources, heavy computation and accessible methods. Therefore, they have obvious needs of annotation standards and ontologies that will make their data accessible to the community and will enable the inter-operation with other sources of knowledge.

- *Projects in specific domains,* for example tree genomics, fish genomics, endangered species or even with greater focus Lynx or Salmon genome projects. Their potential applicability make these projects to be closer to the direct interest of some countries or entities, and in many cases to have a strong industrial support, as well as a rich network of public-private collaborations. Given their dimension these projects tend to have very specific needs, such as access to computing resources, connection to other similar projects, and project/science driven analysis approaches.



- *Specific areas of bioinformatics.* Scientific communities in areas of bioinformatics such as structure prediction, docking, text mining, small RNAs, NGS, transcriptomics, integrated biology and others have also needs of support and use of ELIXIR infrastructure and services. In this case the collaboration of ELIXIR can have an additional beneficial feedback, since the methods developed by these communities contribute to building the portfolio of ELIXIR methods and expertise. These communities have different levels of organization, from formal to completely informal, but in most cases “community challenges” is an important instrument to structure them, to set up goals and define progress (evaluate methods). These activities are important for ELIXIR, since their activity is directly related with the development of standards (see the many publication of communities proposing standards and ontologies).

- *Collaboration with other platforms.* Close collaboration with platforms outside the European Area is envisioned in order to extend the geographical reach of ELIXIR and in synergising with domain specific initiatives. One example is current EMBL Australia mirroring of EBI’s services that is expect to expand to include ELIXIR services and make them available in the Southern Pacific. Another example is the Human Variome Project that is dealing with similar issues as ELIXIR in organizing and sharing of human genetic data and where some concerted action is envisioned.

### **General technical areas of interest across domains**

- *Standards, SOPs and ontologies.*

ELIXIR will collaborate with domain specific communities to develop standards for data organization, transmission, processing and analysis. This work will be more effective if it aims to generalize, and to promote linkage between areas and interrelation between the corresponding data sets and analysis strategies.

The work in SOPs and ontologies is an essential aspect of this activity. Examples in this case can be ontologies for diseases and symptoms and those for the description of metagenomic datasets, including reference databases, descriptions of environments or ontologies driven by the needs of marine biology. Other examples could be the descriptions of samples and experimental approaches in metabolomics initiatives, or needs in standardised descriptions in the various fields dealing with images. In all these cases, ELIXIR can contribute effectively by bringing methodology and expertise gained in the parallel work in different domains.

- *Community challenges and selection of computational approaches.*

ELIXIR can play an additional key role in support of domain specific communities by providing the infrastructure for the development of community challenges. Community efforts such as CASP, CAPRI, BioCreative, and eGASP have turned out to be a natural way of comparing approaches, building resources that can be used by others to evaluate their own methods, agreeing on evaluation standards and, very importantly strengthen links in the respective communities. Given the importance and increasing complexity of these challenges it is clearly possible for ELIXIR to make a positive contribution by promoting good practices, methodologies and standards.

- *Data access, security and interoperability.*

ELIXIR can play an important role in coordinating and implementing systems for providing regulated access to data, ensuring security of data to be kept private, and for making sure data and tools are interoperable. Security of data is of high relevance for sensitive data, for example data relating to individual genomes and phenotype data in context of human



biobanks. Also projects with industrial involvement can have high requirements for data protection and privacy, and the availability of private, secure data analysis frameworks within ELIXIR is critical. For sensitive research data it may be relevant to create meta-databases providing information about available data without exposing sensitive information. This will be of interest for example for medical biobanks.

### **Modes of collaboration, ELIXIR decision process and participation in grants.**

*Bottom up approaches.* For ELIXIR, bottom up is preferable since it will indicate the actual needs of the communities. In practice, it implies a considerable effort to adapt the specificities of each community and therefore requires resources that tend to be underestimated.

*Top down approaches.* In practice ELIXIR should be prepared to adopt a leading role in some areas, helping to put together proposals in direct contact with leaders in some fields.

Overall, the collaboration with a diversity of communities will put ELIXIR in an ideal position to provide common solutions and to develop general strategies, harmonise efforts, share resources, avoid duplication/redundancy and to maximise effectiveness.

*ELIXIR decision process.* For ELIXIR it is essential to develop a process for the selection of communities and projects. Five criteria seems to be generally acceptable: a) Scientific interest, b) availability of resources and c) expertise, d) national commitments and e) relation with other on-going ELIXIR activities.

The HoN and the SAB should be involved in the decision process.

In some cases framework agreements will be recommendable e.g. IMI – ESFRI, in other cases groups of nodes working together in collaboration with experts in specific problem areas might be preferable and more operative, e.g. tree and fish genomics.

*Importance of communication and open information, importance of synergies and operations across domains.*

(To be written.)

### **Challenges in the provision of domain specific services**

We identify four major challenges:

- 1) *to identify active communities and build positive modes of collaboration with them.*
- 2) *to collaborate, support and/or help to establish large scale efforts, like for example IMI projects and BS ESFRI infrastructures*
- 3) *to develop an economically feasible and sustainable model for the work of ELIXIR in specific projects and/or with specific communities*
- 4) *to identify and implement methodologies that can provide solutions applicable across the field, gain from synergies and pave the way to stable infrastructures.*

The alignment with the core bioinformatics resources and the ELIXIR core methods and systems will be essential.



## **ELIXIR objectives 2014-2018 (M1-M6o)**

### *First year*

1. Definition of working models and process and identification of initial potential collaborations.
2. Define a first set of associations of ELIXIR and ELIXIR nodes with specific communities
3. Identify specific aspects/protocols of data handling and security that could be developed by ELIXIR to serve needs across fields.
4. Identify aspects/protocols of ontologies and vocabularies that could be developed by ELIXIR to serve needs across fields.
5. Identify bioinformatics methods and associated communities, which ELIXIR could collaborate with to develop generally applicable validation strategies for important methods.

### *Second year and third year*

6. Initial implementations in specific domains / areas / projects.
7. Set up of formal collaborations.
8. Organization of the operations to retrofit ELIXIR with the expertise, tools and approaches developed in specific domains.
9. Development of methods, resources and standards across domains.

### *Fourth year.*

10. Application of general solutions and strategies to specific domains.
11. Analysis of success and failures according to the community feedback, with possible reorganization of activities.

### *Fifth year.*

12. Consolidation of activities.
13. Implementation of sustainability programs.

## **Resources**

(To be defined.)