ELIXIR-GR SAC report

Comments to the on-line presentations that took place the 25-26th May 2021.

1. **Platforms update: Tools.**

A number of tools were presented as part of the Node and ELIXIR-GR portfolio, including among others: Diana and DIANA-miRPath v3 for the organization and analysis of miRNAs, Proteosign web server for protein differential expression, BioTextQuest a knowledge integration platform for literature mining and concept discovery, metaseqR a tool for the analysis of RNA-Seq data and POTAMOS a mass spectrometry calculator. The implementations in the Galaxy ecosystem were also presented, in particular in terms of analysis of COVID-19 related data.
The individual developments are interesting, have been published and have reasonable user bases. Still, there is a clear need of setting criteria and standards to assess their potential impact as part of the ELIXIR-GR infrastructure. Overall, a better integration with the elements of the ELIXIR tools platform (the bio.tools catalog, the work on BioContainers, the registry of workflows, the tool benchmarking effort and the Galaxy context) will be desirable for the next period of activities.

2. **Platforms update: Databases.**

A catalogue of 15 databases was presented as part of the Node and ELIXIR-GR grant portfolio, covering the field from genomics, proteomics, text mining, and structural variation among others.
The highlights were Diana mirR V4 with a large user base, associated with a reasonable processing capacity that have gained a well-deserved international reputation in this competitive area. Diana-LncBase V2, with information on experimental and predicted interactions between miRNA and lncRNA, is also well published with many citations and now is connected to ENSEMBL. Pickle is a human protein-protein interaction metadatabase employing primary dataset integration via genetic information ontology with new features for the interpretation of mouse-human data sets. HALODOM provides an organisation point for research into halophiles with data collections and tools of relevance. BIP4COVID19 is an implementation of the BIP search engine for the analysis of publications related to Covid-19. All in all, interesting developments that are supported at a relatively low level in comparison with their impact and number of users.
Once the resources are established for the next phase ELIXIR-GR should consider following the ELIXIR recommendations for Core resources, in a way that the different databases can be categorised by the level of fulfilment of those criteria. Even if there is no need for any database to be a core database, or if participation in this part of ELIXIR might be as one of several clustered resources (such as is the case for InterPro), this exercise will contribute to clarify their status, define uniqueness and offer future perspectives. It is important to highlight that ELIXIR-GR would support the application of the DIANA TarBase as an ELIXIR Core Data Resource, for what the University of Thessaly would have to join the ELIXIR-GR consortium agreement. The selection of core databases is a very demanding category with a number of specific criteria that will be evaluated by external referees as well as by the ELIXIR Head of Node. The SAB encourage all the involved parties to make this possible, since – if successful, the ELIXIR endorsement will contribute to the long-term sustainability of this resource.
More broadly, we encourage engagement with the ELIXIR Data Platform, where ELIXIR-GR may become part of the ELIXIR Core Data Resource discussions, other existing areas of the Platform's work (such as on the publications side) or as a proponent of new areas of work.

3. **Platforms update: Training.**

The training activities delivered and programmed for the period 2020-21 were presented.
During this period ELIXIR-GR has been a very active member of the ELIXIR training platform. Indeed, Fotis Psomopoulos, is one of the coordinators of the ELIXIR training platform together with Jessica Lindvall and Celia van Gelder, as he is also the leader of the task force on “Gap analysis, training

materials development and training delivery". In the historical series, the total number of training activities since 2018 account for 40 days of direct training and about 600 participants since 2018.
The dissemination of the activities seems to be done mainly by email, while the ELIXIR-GR website dedicated to the training platform seems to need an injection of information about past and planned events, with dates, venues, topics and links. The names and responsibilities of coordinators and members of the team working in the organization of training activities should also be available for contact and planning of future activities. While there is a recognised effort in this area, the complex times ahead will require additional efforts to concentrate and organise the ELIXIR-GR training capacity in collaboration with the local universities in the framework of the ELIXIR activities and guidelines.

4. **Platforms update: Compute.**

It is unfortunate that the ELIXIR-GR infrastructure has been delayed in its deployment both for Corona-dependent reason and by a potential lack of understanding that a computational infrastructure is essential to the computational biology/bioinformatics of the 21st century. It is obvious that relying solely on local infrastructure and the provisional services provided by GRN-net was insufficient to provide nation-wide infrastructure and was not sustainable in the long run.
The SAC is happy to see that the compute infrastructure, Hypathia, with 32 nodes and 2 fat nodes, seems to be now in the final phase for entering in full operation (61 users with 33 projects were mentioned), while issues of access, allocation of resources and management of virtual machines have still to be fully solved.
With the new ELIXIR-GR central infrastructure, Greece can build through GR-Net (on the network side) a resilient infrastructure. The description of the current interface for the non-bioinformatician has been well explained and can reach potentially a good user base.
The ELIXIR-GR Galaxy instance (https://usegalaxy.elixir-greece.org) was presented as operational and open for registration, linking ELIXIR-GR with the large ELIXIR Galaxy community.

5. **Pilot actions update: Marine metagenomics.**

The Marine metagenomics pilot corresponds to one of the landmark activities of Greek science, therefore, it is considered a key activity for ELIXIR-GR. The reported activities included the establishment of the pipelines for Nanopore and Illumina data and for the analysis of microRNAs. Among the interesting developments the collaboration with LIfeWatch Greece and Bioimage for the development of methods for image analysis was mentioned, as well as the possible participation in ERGA. This collaboration will connect SAC report 2021 ELIXIR-GR with the ELIXIR biodiversity community and can be considered strategically important to position ELIXIR-GR for the Horizon2020 new round of grants.
The SAC very much supports the efforts of ELIXIR-GR in this area, appreciate the contribution of the institutions, in particular HCMR-IMBCC that is providing the compute capacity and resources, and would encourage all parties to continue to work improving the internal coordination, as well as the linkage with the other ELIXIR nodes and European activities.

6. **Pilot actions update: Metabolomics & protein interactomics.**

The pilot activity bridging metabolomics and protein interaction is developing well and has acquired a considerable international profile. ELIXIR-GR has been involved in the creation of the ELIXIR metabolomic community and is part of the "Standardising the fluxomics workflows" commissioned service.
It is likely that this project will be more fulfilled in the area of Agritech and Greentech through the metabolic engineering side. Orientation with the marine metagenomic is an obvious step which has been in part taken by the participants of ELIXIR-GR. We are looking forward to the development and enhancement of this pilot project which piggyback on already existing resources and contribute with novel development within Greece.

7 & 8. Pilot actions update: Noncoding RNA biomarkers and Pathogen metagenomics.

These two related actions account for a number of related topics covering human and pathogen omics, ncRNA (miRNA, lncRNA), verified circulating miRNA biomarkers with applications in cancer and covid-19 use cases. At the technical level the work described covers aspects from data collection and databasing to method development (i.e. ML was introduced), causal networks and genetic associations, as well as specific implementations (Galaxy was mentioned). With applications ranging from biomarker detection to disease prognosis.

In the case of cancer, the topics included pipelines for the analysis of microRNA and lncRNAs and their potential interactions with applications to data from different sources in at least two cancer types: glioblastoma and breast cancer. The overall level of activity, accompanied with international collaborations (In general, U. of Pennsylvania, Luxembourg Institute of Health, the U of Athens were mentioned), as well as with local clinicians, with a number of relevant publications is quite impressive and demonstrates not only the utility of the tools (i.e. TARBASE resource) but also the expertise associated with them.

For future developments, the SAC would have liked to see a clearer relation between the pilots and the developments at the level of ELIXIR-GR, i.e. how this work becomes more than a number of research projects and contributes to the development of the overall ELIXIR infrastructure.

9. Evolving Platforms: Interoperability.

ELIXIR-GR is increasingly visible in the context of the general ELIXIR activities such as FairSharing, ELIXIR AAI services and workflow management particularly in relation to COVID data.

As a general consideration, interoperability, while being an important task, is too much of a placeholder, not only for ELIXIR-GR but for ELIXIR itself and for all ESFRI. There is a need to interoperate resources and two choices are possible, ELIXIR-GR should identify its own, or contribute to already existing resource through curation, data stewardship, standard development or to develop its own interoperability within the ELIXIR-GR resources (e.g. marine biology which is a highlight of Greek developments).

The evolution of the platform is becoming complicated by the fact that Greece is seeking a Core Resource within its portfolio, the interoperability is the critical part which creates interdependencies. Seeking these interdependencies will render a resource a Core Resource, regardless of the number of users.

10. Management of Human Data.

The presentation describes how ELIXIR-GR has adhered to the ELIXIR FAIR data deposition standards including the deposition of human data to the public EGA databases, an effort accompanied by the work in training users and developers and motivating data sharing behaviours. Along this line the efforts to install a local EGA repository should be completed, and its maintenance supported, since it will represent the connectivity of Greece with European initiatives such as the 1MGenomes.

The SAC encourages the efforts of ELIXIR-GR as well to collaborate with the Rare Diseases and Health Data/Biomedical Engineering local communities and to connect the work with the corresponding ELIXIR communities and the Health Data Focus Group.

The alignment of the work on human genomics with the investigation of human microbiome diversity and connection with diseases was also an interesting focus point that gives coherence to the work of a number of ELIXIR-GR teams.

Overall, for the following period of activity it will be good to see a consolidation of the efforts around the EGA database, addressing the issues of human data storage but also the ones of data discovery (beacons), mining and analysis, linking the ongoing ELIXIR-GR activities (e.g. modelling of human metabolisms) as the basis for a more substantial contribution to the global ELIXIR efforts.

11. COVID-19 activities.

ELIXIR-GR has made a considerable effort to provide services in the context of the pandemic, aligning with general ELIXIR work. In particular, the contributions are notable to establishing the national COVID-19 portal, including the Galaxy pipelines for the analysis, rallying for the mobilisation of sequencing and deposition of viral genomes, engaging in the public health discussion around open data and brokering viral sequence to ENA. The issues with submission to the open data EGA repository, aligning with the current ELIXIR efforts, were discussed and encouraged in this context.
The SAC also encouraged ELIXIR-GR to contribute to the efforts of lead by the Disease Map community to build maps of viral-host as part of a large international effort, to which ELIXIR-GR can bring the expertise in protein interaction analysis, mining related publications and/or developing approaches closer to proteomic and modelling.

12. Community/Focus Group engagement

ELIXIR-GR participates in almost all the ELIXIR focus communities, particularly Marine Metagenomics, Metabolomics and Microbial Biotechnology, with intentions to have lead positions in Disorder Proteins, Metabolomics, Microbiology and Rare Diseases. In the case of the communities, there is also a high degree of involvement, in the Machine Learning, Fair Training and Biodiversity ones.
In the opinion of the SAC these activities, even if costly in terms of time and resources, are very beneficial for shaping the internal structure and developments lines of ELIXIR-GR, for fostering the adoption of guidelines and procedures as contributors and not as mere spectators and to make visible the work done by the Greek bioinformaticians.

13. The future of ELIXIR-GR.

The presentation of the strategic plan for the years 2022-2028 attends a request of the SAC and fulfils an important strategical need for the consolidation of the ELIXIR-GR activities. The presentation of the plan was clear and comprehensive.
The SAC agrees and encourages the plans to further integrate into the overall ELIXIR structure and organization. A lot of progress has been made since the previous SAC visit and we note that there is substantially greater awareness and engagement across ELIXIR since the previous sitting of the SAC; still there are areas of the organization of ELIXIR-GR that will benefit from following and engaging in ELIXIR developments. The definition of standards for databases and the associated criteria for core databases is a good example of areas that will require further attention.
The description made of the key developments in Data, Tools, compute, training and interoperability coincides very well with the opinion of the SAC expressed in this document.
A particular point of attention was the coordination of the work done in Cyprus in the activities of ELIXIR-GR. Given the recognised quality of the work done by the bioinformatics groups in Cyprus, the SAC very much supports this proposed collaboration.

SAC conclusions and recommendations.

The conclusions of the virtual SAC, as initially summarised during the meeting, are the following:

1) First, we congratulate ELIXIR-GR for the smooth transition in leadership and for the perceived positive work environment created since the previous visit.
2) The SAC very much appreciates the work done to adapt to the structure of ELIXIR, particularly in this difficult COVID period.
3) On the logistic side, the program and documentation were well organised following the ELIXIR areas of activity and sufficient for the online evaluation. Still, for future occasions it will be good to have a clearer overview of the impact of the tools, involvement in the ELIXIR activities, distribution of resources and decision processes. All these aspects that have been vastly improved but can always be further refined.
4) The SAC values the international standing of a number of resources hosted by ELIXIR-GR in fields of genomics, metabolomics, expression (miRNAs in particular), marine metagenomics,

microbiology, as well as the accompanying technical developments in different areas, e.g. text mining.

5) The recent developments to finally establish a common compute infrastructure and the steps to make it operational and accessible are considered a very important and fundamental point for the future.
6) The contribution to the Galaxy project, the involvement with the ELIXIR Galaxy community and the associated developments (e.g. analysis of SARS-COV-2 genomes) are very positive as initial steps towards the establishment of Galaxy as a national analysis platform.
7) The work in the area of interoperability is maturing to the point that it will be important for ELIXIR-GR to decide how to adapt them to its own needs. A process that will require concentration of resources and decisions about future implementations aligned with the data and compute resources.
8) Along this line, curation of data associated with some of the resources is an area in which ELIXIR-GR should look into since it can provide the differential factor that some of the current resources may need to have a global impact.
9) The initiative to develop pilot actions is considered valuable and successful for this phase of the development of ELIXIR-GR since they have contributed to speed up developments. For the future, if additional pilots are implemented, it will be necessary to link them better to the development of the ELIXIR-GR infrastructure.
10) The SAC looks forward to learning about progress in the training side where a more determined effort in concentration and organization will be required - as well as adaptation to the constraints imposed by the pandemic.
11) Finally, the SAC is very happy to see that ELIXIR-GR is now aligned with ELIXIR platforms, communities and focus groups. It is also good to see that ELIXIR-GR is progressively taking positions of responsibility in some of these areas. All of which makes by now the organization of ELIXIR-GR comparable to any of the best organised ELIXIR countries.
12) Based on these positive developments, the SAC strongly urges the Greek authorities to support and finance appropriately the activities of ELIXIR-GR since, by bringing both uniqueness and additional strength to global life science infrastructure, they are essential for the development of biology and biomedical science and industry in Greece and Europe.

SAC Members

Guy Cochrane, EMBL European Bioinformatics Institute (EMBL-EBI)
Manuela Helmer Citterich, University of Rome Tor Vergata
Alfonso Valencia, Barcelona Supercomputing Center, BSC-CNS
Ioannis Xenarios, Université de Lausanne
(the additional SAC member; Nikos Kyrpides, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, was absent with apologies)



Guy Cochrane
EMBL European Bioinformatics Institute (EMBL-EBI)

Manuela Helmer Citterich
University of Tor Vergata

Alfonso Valencia
BSC-CNS