EMBL

**EMBL-EBI Services Review**
**26 - 28 March 2019**

EMBL-EBI Services were reviewed on 26 – 28 March 2019 by a panel of 18 international experts, including four members of the Scientific Advisory Committee (SAC). The review was chaired by Edward Marcotte from the Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, Austin (US). The Chair of SAC Paul Nurse attended the review as an observer.

**Evaluation Summary**

By all metrics, EMBL-EBI services are a success, vital in delivering biological data world-wide. The services at EMBL-EBI appear more cohesive and coherent than ever before, particularly in the context of the explosion of data which has occurred since the last review in 2015. This is in part due to the structuring of the service teams into thematic clusters under Rolf Apweiler and Ewan Birney's leadership. This has improved the coordination and integration between related EMBL-EBI service areas and facilitated user access to resources. Thanks to this exceptional leadership, EMBL-EBI is placed at the forefront of many of the most important data initiatives worldwide, such as the Global Alliance for Genomics and Health (GA4GH), the Human Cell Atlas (HCA) and ELIXIR, within which 13 EMBL-EBI resources are selected as Core Data Resources in Europe. EMBL-EBI websites serve tens of millions of web requests every day and they provide biological and biomolecular data, services, and infrastructure to the academic, clinical and industrial life science communities. This includes archiving, curating, and integrating data, which span the chemical, molecular, cellular, phenotypic and literature realms.

Several notable achievements were highlighted by the panel, including the successful deployment of the Open Targets Platform in 2015. The panel felt that Open Targets was an important driver of integration across multiple cluster resources and encouraged the unit to consider how to take advantage of this in other domains that are of interest to the user community. The review panel was enthusiastic about EMBL-EBI's growing efforts in the Global Biodata Coalition whose aim is to coordinate long-term funding for core resources and strongly supports further efforts in this area.

The panel also highlighted the development of the Electron Microscopy Data Bank (EMDB) and the associated Electron Microscopy Public Image Archive (EMPIAR), both of which are having an enormous impact in the cryo-electron microscopy field. Additionally, the panel commended the remarkable growth of the European Nucleotide Archive (ENA), EMBL-EBI's main archive for primary nucleotide sequence data and derived analyses that support many critical downstream services. Alongside this, the proteomics identifications (PRIDE) database has become the leading proteomics data repository in the world and has transformed the field of proteomics by being the dominant driver of open data culture.

EMBL-EBI is commended for expanding into biodiversity and ecology genomics, with a large network of partners in place around the world in various tree of life sequencing projects. In this context, the panel commends EMBL-EBI for embracing the Nagoya protocol, actively participating in talks about ethical hosting of data, and developing pilot projects around data provenance tracking. The panel recommends that EMBL-EBI take a leadership role in promoting good practices for considering ethical and moral aspects of sequencing datasets and analyses.

The panel felt that the plan to develop a bioimage database was exciting and timely. It will need to start with a clear remit, and it was recommended that a new staff member move this forward, together with Ardan Patwardhan and Alvis Brazma, as proposed. There was enthusiastic support for the mechanism by which ArrayExpress, within the Molecular Atlas Cluster, was phased out as a major resource. The panel recommended that this effort should be taken as a basis for a procedure for future decommissioning decisions. The panel also noted that consideration be given as to how to strengthen this cluster and sustain key resources, given the small size of the cluster and potential for destabilisation. This could be applicable to the context of rapid developments in data science at large and of which EMBL-EBI needs to take note. The panel recommended that EMBL-EBI set up a small service unit which will provide data science expertise across the institution.

With the continuous growth of data and of data types in life sciences, it is difficult to attract all necessary expertise in bioinformatics at the current EMBL-EBI site. The panel recommended that this calls for some distribution of data resources across other European sites. ELIXIR appears to be the natural ecosystem for these to emerge and consolidate, but some new service areas could be directly deployed by EMBL-EBI away from the main site. This will engage the Member States more strongly and facilitate investments into infrastructure by other European governments, similar to those by the UK government.

Moving forward, EMBL-EBI's plans for the next four-year period appear to be appropriate, with a strong emphasis on human health and increasing efforts on agriculture, biotechnology, and the environment. The panel is pleased to note this is planned to be done in the context of building stronger connections to other EMBL units, industry, and international research consortia.

**Response to the Panel's Recommendations**

I would like to extend my thanks to the panel for their time and effort in reviewing EMBL-EBI's service activities. It is wonderful to hear the exceptionally positive report of the services provided by EMBL-EBI and I echo the panel's commendation of EMBL-EBI's leadership. I would like to congratulate Ewan Birney and Rolf Apweiler for creating a cohesive and well-structured working environment that provides world-leading data services to users in academic, clinical and industry area alike. The excellent integration and interconnectivity of services within EMBL-EBI further builds on EMBL's pre-existing collaborative strengths. In this context, I am pleased to recognise the success of Open Targets as a driver in integrating multiple cluster resources. The panel recommended applying this successful model to other domains. The very substantial ground work and manpower that would need to be deployed in launching another such operation would need to be considered carefully.

As recommended by the panel, I would like to note that the Global Biodata Coalition is a good opportunity to coordinate long-term funding for core resources. However, I am aware that shifting external funding landscapes means we cannot depend on this and EMBL is highly appreciative of vital funding given by its member states. In line with EMBL's service mission, this core funding drives these data services, without which researchers and clinicians in the member states and beyond would be at a complete loss.

I would also like to highlight that in his capacity as the leader of the GA4GH consortium, Ewan Birney is uniformly trusted around the world as an expert in health data and he has my full support in this endeavour.

The recognition of EMBL-EBI's expansion into biodiversity and ecology genetics is greatly appreciated. I would like to take this opportunity to highlight the research and service links this forms with other areas of EMBL which explore the molecular mechanisms behind biodiverse organisms, in particular EMBL's collaboration with Tara Oceans. As part of this hugely successful global project, EMBL-EBI services have been integral to making unique and novel oceanic genomic data globally accessible. Across units, EMBL is also actively discussing how to further integrate these topics into EMBL's activities across all of the missions. I would like to praise EMBL-EBI for its proactive and responsible approach to the ethical hosting of data and I fully support the recommendation that EMBL-EBI take a leadership role in promoting good practices for considering ethical and moral aspects of sequencing datasets and analyses. EMBL-EBI should also be congratulated for setting global standards as an open access provider for exceptionally high-quality data.

I agree with the recommendation that it is vital to maintain a broad and full understanding of the data services required by the community and to respond accordingly. This is exemplified by the elegant decommissioning of ArrayExpress and the ongoing discussions regarding the setting up of the BioImage Archive. The BioImage Archive is an ambitious initiative that would further integrate current and future imaging services and infrastructures across many of EMBL's sites. However, the funding and personnel required to make this a success must be carefully considered within the context of current funding for the EMBL Programme.

The panel's recommendation to consider the rapid developments in data science and setting up of data science institutes was also noted. EMBL-EBI and other EMBL sites are already actively involved in using AI and other approaches. Indeed, all EMBL would like to move towards powerful data science approaches in the future and this topic will be considered more generally in the context of the preparation of the next EMBL Programme and Indicative Scheme.

EMBL-EBI has grown both because of, and in response to, its continued success. As biology becomes more data driven and dependent on informatics, the service component of EMBL-EBI will need to be further developed. The recent £44.5M funding awarded by the UKRI's Strategic Priority Fund will enable the substantial increase in EMBL-EBI infrastructure that is necessary to meet the rapidly increasing demands made on its services. There are also plans for a potential new building on the Wellcome Genome Campus. I am supportive of this on-site growth for EMBL-EBI, as part of EMBL's service mission. I will maintain an awareness of EMBL-EBI's continued expansion in the context of funding for the next EMBL Programme. The panel suggested that some new service areas might be directly deployed by EMBL-EBI off their main site and within other member states. It is my opinion that the distribution of data resources should be limited to the infrastructure provided by ELIXIR, nevertheless EMBL-EBI will continue to collaborate and advise actively across Europe.

To summarise, I would like to congratulate EMBL-EBI services for their performance over the last four years and credit Ewan Birney and Rolf Apweiler for their cohesive and exceptional leadership.

**Professor Edith Heard, FRS**
**Director General**

05 June 2019