

## EMBL-EBI Services Review Summary and Response

The review of the EMBL-EBI Services took place March 21-23, 2023, at EMBL-EBI, Hinxton, United Kingdom. The review panel comprised 30 external reviewers, 5 members of EMBL's Scientific Advisory Committee (SAC), and additional internal observers. The review was held in a mainly virtual format due to the current financial restrictions, with only the Review Panel Chair and Sub-Chairs, three members of the Scientific Advisory Committee, and the EMBL observers including the EMBL Director General, attending in person.

The review was chaired by Edward Marcotte from University of Texas at Austin and the sub-chairs were Caroline Relton from the University of Bristol and Bas van Steensel from the Netherlands Cancer Institute. Observers included EMBL Council Chair Peter Becker, EMBL Council Vice Chair Amanda Collis, EMBL Director General Edith Heard, Joint Director of EMBL-EBI Rolf Apweiler, Joint Director of EMBL-EBI and EMBL Deputy Director Ewan Birney, and Chair of SAC James Briscoe.

### Evaluation Summary

The panel highly commended the quality of EMBL-EBI Services and noted that its service role is unique and unmatched. Furthermore, EMBL-EBI combines dedication to open science and interconnectivity with cutting-edge biological research. Over the past four years, it has risen to become arguably the leading bioinformatics service-provision institute worldwide. Since the last review in March 2019, EMBL-EBI has produced many new scientific services, not least the essential contribution of the COVID-19 data portal in the fight against SARS-CoV-2, as well as the implementation of deep learning advances, such as in protein structure prediction with AlphaFoldDB.

Several leadership and structural changes have taken place at EMBL-EBI since the last review. The organisation made a transition from two Associate Directors for EMBL-EBI Services to one and redesigned the structure to feature four large sections in place of eight thematic units of uneven size. The panel commented that the transitions went well, and noted particularly that this success was achieved when much of the organisation was working remotely.

During the review period, EMBL-EBI has continued to offer world-leading services and launched five new services, including the Bioimage Archive, AlphaFoldDB, the COVID-19 Data Portal, Federated EGA, and DECIPHER, which was transitioned from the Wellcome Sanger Institute to EMBL-EBI in 2022. The importance of EMBL-EBI's continuing services was enunciated with a hypothetical deletion test: the panel considered the impact if a service such as Ensembl, UniProt, or ChEMBL were stopped tomorrow. They concluded that there are no suitable alternatives worldwide that provide equivalent services, and researchers around the world would be significantly negatively impacted. The panel added that the uniqueness of EMBL-EBI resources

also stems from their interconnectedness and cross-integration, providing additional power to the services and allowing for significant internal efficiencies.

While noting that EMBL-EBI's prodigious output provides many examples of scientific highlights, the panel selected three to mention specifically. First, EMBL-EBI's response to COVID-19 was remarkable for its rapid deployment of COVID data portals, which served important roles in guiding the European response to the pandemic and provided the foundation for the European Commission's COVID-19 platform. Second, the development of AlphaFoldDB (which provides computed 3D structures for essentially all protein sequences in the UniProt database) had a broad impact on global research by offering open accessibility to protein structures. Third, the launch across six countries of the Federated EGA, a working federated model for sharing controlled-access human data, lays the foundation for future efforts to responsibly and easily share access to this important research data.

During the review, EMBL-EBI's future plans emphasised opportunities to leverage advances in AI and data sciences. The panel agreed that EMBL-EBI is uniquely positioned to bridge efforts to apply AI to biological datasets, provide high-quality reference data for training AI, and vet and disseminate the results to user communities. The recent examples of such ventures are highly successful and include the DeepMind partnership to incorporate AlphaFold structures into UniProt, and a partnership with Google on domain analysis. These initial projects point to significant promise for future work, and the panel recommended EMBL-EBI continue to grow in this area.

Overall, the panel noted high integration with EMBL-wide activities and research. EMBL-EBI Services activities are well-aligned with EMBL's new "Molecules to Ecosystems" programme. The level of integration between services and research, both internally at EMBL-EBI and across EMBL, has improved significantly over the past four years.

The panel had group meetings with content staff, technical staff, and service coordinators and leads during the three-day review. In general, staff were uniformly positive about the sense of community and pride at EMBL-EBI, and noted particularly the response to the COVID-19 pandemic and the current flexibility in work environment that has persisted. Staff representatives suggested that better communication and more coordination between services would ease some of the challenges around implementation of changes and migration.

Staff also raised concerns about the continuity of services given EMBL's nine-year turnover model, and noted that difficulty in recruiting for technical roles exacerbated the inherent challenges in the turnover model. Staff also noted career advancement and professional development, both internally and externally, as areas of improvement.

The panel made several recommendations for EMBL-EBI Services to retain its premier status globally and to continue to improve its service offerings.

Given the financial challenges that EMBL currently faces due to unprecedented inflation and energy costs, funding was the top issue of concern. The panel recommended that EMBL-EBI Services explore philanthropic funding sources and/or increase project-based grants for appropriate activities. They also noted that EMBL-EBI genomic resources have a large impact in animal and plant breeding, and better links to AgTech may open new avenues for funding. Funding of EVA beyond 2023 was an additional concern. However, the panel noted that this resource, as well as many other EMBL-EBI services, are recognised as core resources by ELIXIR and the Global Biodata Coalition and the panel looks forward to seeing how this recognition translates into stable funding. Finally, the panel suggested expanding on systematic ways to evaluate EMBL-EBI resources in order to identify those for which the investment in maintenance outweighs the active use value and which are candidates for retirement.

The panel made two specific recommendations regarding team leadership. They noted that it is of high priority to fill the gap that would provide a scientific link between the human genetics research community and relevant EMBL-EBI Services. They also noted that team leaders seem to have many and diverse responsibilities, and suggested that simplifying their responsibilities may help focus their activities.

EMBL-EBI Services were commended for improvements made in the diversity of personnel, especially at higher management levels; however the panel noted that there is still room for improvement in the area of equality, diversity, and inclusion (EDI) principles and practices. They also recommended that career development opportunities for staff should be improved, particularly for technical staff. The provision of leadership training for senior staff and mentorship opportunities at all career levels should also be increased. The panel commended EMBL-EBI for maintaining services even with the high staff turnover, and for its detailed tracking of staff leavers. The panel noted that the latter documents the downstream positive impact of EMBL-EBI staff training on the greater Data for Life Sciences landscape in Europe and worldwide.

In conclusion, the panel found Johanna McEntyre's leadership of EMBL-EBI Services, with the support of EMBL-EBI co-directors Rolf Apweiler and Ewan Birney, has been exceptional during the review period. The panel reiterated that EMBL-EBI services are the best in the world and emphasised the need for them to be supported to the maximum extent possible.

## Response to the Panel's Recommendations

I would first like to extend my sincere thanks to the panel chair, Edward Marcotte, the panel sub-chairs Caroline Relton and Bas van Steensel, and the entire panel. I appreciate the time and dedication that they put into the review preparation, and the depth of knowledge, understanding, and careful consideration that they brought to the three-day discussion and the written report. I share the panel's enthusiasm and belief that EMBL-EBI services are uniquely world-class and

irreplaceable. I would also like to congratulate the EMBL-EBI Services Director, Johanna McEntyre, for her remarkable leadership over the past four years, particularly during a global pandemic that was challenging for all and during which EMBL-EBI Services played a critical role by continuing to provide its services to the world. I am also very pleased to see that the efforts to integrate EMBL-EBI services more closely into EMBL's research activities across its multiple sites are recognized and noted, particularly the role that the EMBL-EBI staff have played in realising and contributing to the 2022-2026 EMBL Programme, "Molecules to Ecosystems".

EMBL-EBI resources are unparalleled and are of ever-increasing importance to life sciences research, not only in academia, but also in health care, the agricultural sector, and industry. I therefore share the panel's concerns regarding sustainable, long-term funding. Funding is an ongoing concern not only for maintenance of core resources, but for the long-term sustainability of growing data resources and growing infrastructure needs. We hope that through ELIXIR and the Global Biodata Consortium we can also demonstrate to national ministries that services and databases such as those offered by EMBL-EBI require long-term, sustainable funding in place of short-term grants. I also appreciate the panel's suggestions to pursue philanthropic funding, project-based grants, and AgTech industry relationships more extensively, and we will continue our efforts in these and other short-term funding options.

EMBL's 9-year turnover model inherently puts significant pressure on services, particularly in continuity of knowledge and recruitment efforts. I am pleased that EMBL-EBI's strategies for maintaining services with high staff turnover are positively noted by the panel, as is the value of the 9-year rule in positively impacting the broad life science landscape globally. That said, I recognize that when recruitment is an issue and longstanding staff members leave, it is not only challenging to maintain services, but morale can also suffer. We will continue to work to ensure that EMBL's turnover model does not become a hindrance, and will strive to ensure that recruitment is more effective.

I thank the panel for highlighting some of EMBL-EBI's Services' gaps in career development. I agree that there is more we can do to support our staff while they are at EMBL-EBI, and better ways to encourage and facilitate their transition to other opportunities. We will take the panel's specific suggestions into consideration, and look into developing clear career pathways and mentorship schemes for personnel whilst they are at EMBL-EBI and for future transitions to other positions.

Concerning the panel's recommendation to ensure better links with the human genetics research community, I agree with the panel, and know that Johanna, Rolf, and Ewan also all strongly agree, that it will be important to find a means to replace the knowledge contributed by Paul Flicek in this area. We assure the panel that filling this gap is a priority and we will look both internally and externally to do so.

Looking to the future, EMBL is very excited about the potential that AI has to disrupt biological data and biological research. We expect that EMBL-EBI data services and its reference datasets will play a critical role in the development and availability of AI technologies for the life sciences. The field is still in its infancy, but we can already note that AlphaFold would not exist were it not for the high quality, open-access data that is available thanks to EMBL-EBI databases including UniProt, MGnify, and PDBe. I concur with the panel's recommendation that EMBL-EBI's widespread embrace of AI and deep learning methods across all services should continue to grow and expand in new directions.

I would like to conclude by once again congratulating Johanna McEntyre and everyone at EMBL-EBI Services, particularly all the Team Leaders, for a very successful review. EMBL is extremely proud of what they have achieved over the review period and I am looking forward to supporting the Unit in the years to come. My thanks to Panel Chair Edward Marcotte, sub-chairs Caroline Relton and Bas van Steensel, and all the review panel for their fruitful discussions, advice, and expert contributions, which will be crucial to the future success of EMBL-EBI Services.



Professor Edith Heard, FRS  
Director General  
27 April 2023