

## **REQUEST FOR SUPPORT FROM THE GOVERNING COUNCIL SPECIAL FUND: SCIENTIFIC IT PLATFORM**

1. Information Technology is a fundamental resource for IARC's scientific activities. This proposal aims to introduce a modern approach to all of IARC's scientific data whether epidemiological, omics, images or databases.
2. Currently, scientific data is stored across multiple systems, which leads to increased maintenance costs, difficulties in moving data, and a lack of a global view of what scientific data is held at IARC. Data analysis is often performed on local computers increasing security concerns about the safety of the data held on those devices.
3. With this proposal, we aim to provide the necessary infrastructure to store all of IARC's Scientific data safely, consistent with current best practices and regulation such as the EU's General Data Protection Regulation (GDPR), consolidating our data centrally to allow efficient and easy access as well as fostering Open Science through data sharing.
4. IARC has already started to implement such an approach for large datasets (mostly omics) requiring high performance computing (HPC). IARC's current HPC environment was put into production in September 2018, partially funded by the Governing Council Special Fund in 2017 and old environments were decommissioned. The migration phase was used to introduce a new data centric model for both storage and management of large data sets, which has proved to be a great success. The Agency progressed from four users of our HPC platform in 2014 to 19 users in 2016 to our current level of 70 users across nine Groups and five Sections and 80 scientific projects storing their data on this system.
5. Applying the lessons learned from the HPC platform, we aim to generalize this approach to all of IARC's scientific data, by providing a unified centralized storage system, including backup and archive. The creation of a catalogue of our data will help to alleviate the lack of a global view of all our scientific data. The provision of associated data analysis software and hardware will ensure that our data stays in one unique place where scientists can work directly on the data with tools such as R, SAS and Stata, as well as potentially allowing remote access to these resources to external collaborators.

6. In line with the governance of the current HPC resources, as outlined in a request presented in 2017 (see document [SC/53/6](#) and Governing Council [Resolution GC/59/R9](#)), the Scientific IT platform would be operated as a shared resource under the responsibility of the IT Working Group of the Computational Biology, Biostatistics and Bioinformatics Committee (C3B), who would provide access and support to the other research groups at IARC.

7. The estimated cost of this proposal for storage servers (€115 000), compute servers (€70 000), network equipment (€25 000), data management and analysis software (€120 000), professional services (€20 000) as well as maintenance of the platform, for a period of four years is €350 000.

8. The Scientific Council is requested to consider this proposal for support from the Governing Council Special Fund and to make its recommendations to the Governing Council.