

## Data Management and Bioinformatics Needs & Skills in Portugal

### 1. Introduction

BioData.pt implemented a national survey to better understand the needs and skills in data management and bioinformatics of the national scientific community, with the goal of adjusting its service offer.

The survey was prepared using a [Google form](#) and sought to balance coverage of the topics with ease of responding, so as to avoid surveyees quitting mid-survey. It was disseminated through direct institutional channels, through a newsletter targeting researchers working in public and private Portuguese institutions, and via social media and the BioData.pt website.

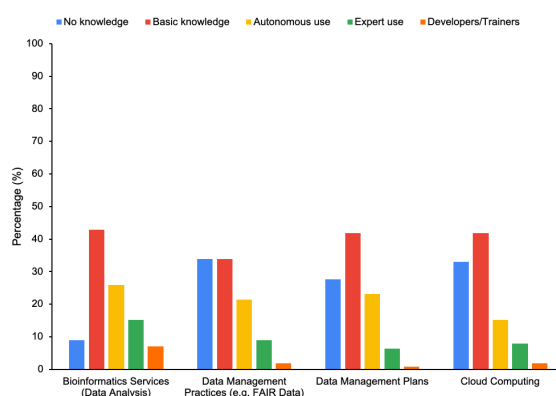
The survey was released on April 17<sup>th</sup> and closed on April 30<sup>th</sup>, 2020.

### 2. Results

The survey was completed by 112 participants from 31 different organizations, 17 of which are located in the Lisbon region, 9 in the North, 2 in the Center, and 3 in the South. The 11 R&D institutions that belong to the BioData.pt Consortium accounted for 69 of the participants.

With regard to affiliation, 60% of the participants hailed from universities or academic institutes, 35% from research institutes, 3% from other public or non-profit institutions, and only 2% from private companies. Concerning their roles, 29% of the participants were senior researchers, 20% were post-doctoral researchers, 24% were PhD students, 10% were lab managers or technicians, and 17% had other roles.

In terms of skills, the general picture is that none of the four surveyed topics is particularly well covered by the surveyees' group (**Figure 1**). Bioinformatics is the best covered topic, with 26% of the surveyees considering their group are "autonomous users", and 22% declaring their group are "experts" or "developers/trainers", but this still leaves

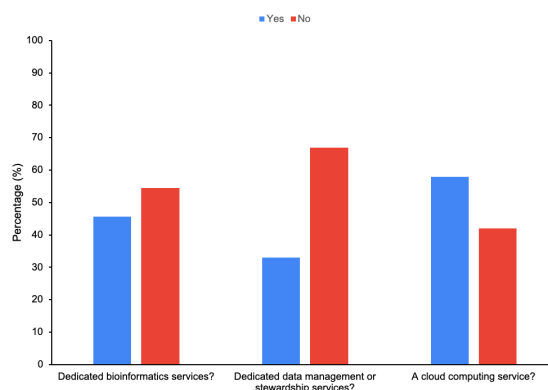


**Figure 1.** Existing skills among the surveyees' groups.

out 52% who consider their group has "basic knowledge" or "no knowledge". The remaining three topics fared worse, with the combination of "no knowledge" and "basic knowledge"

accounting for 68%, 70% and 75% of the surveyees in data management practices, data management plans, and cloud computing, respectively.

With respect to the availability of services in the surveyees' institutions (**Figure 2**), cloud computing services are available to most surveyees, but bioinformatics services and data management or stewardship services are available to only 46% and 33%, respectively.



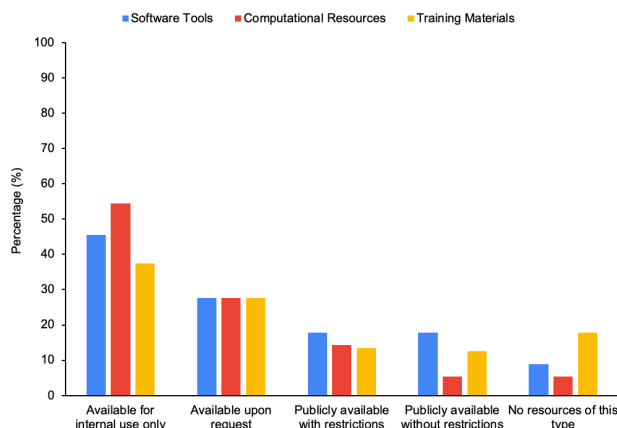
**Figure 2.** Availability of services in the surveyees' institutions.

In terms of needs (**Figure 3**), over 55% of surveyees recognized the need for training in all four surveyed topics, at least 40% recognized the need for user support or consulting in those topics, 15-40% identified the need for some time of resources (tools, computational resources or documentation), and less than 10% declared they had no needs.



**Figure 3.** Types of needs of the surveyees.

Regarding available resources (**Figure 4**), most surveyees declared that they had access to software tools, computational resources and training materials in their groups, but generally these are available for internal use only, or at best available upon request. Some resources are publicly available with restrictions, and few are freely available, particularly with respect to computational resources. Training materials were the type of resource that most surveyees declared lack of access to, in agreement with the previously identified training needs.



**Figure 4.** Resource availability - Software tools, Computational Resources and Training Materials.

Concerning data sharing, 86% of surveyees perceived the benefit of having all research data on their topic of interest integrated and publicly available and integrated, and 97% declared themselves willing to publish their data in public repositories, though most only after the research was already the subject of scientific publication.

### 3. Conclusions

The results of the study show that a clear gap between the needs and skills on the topics of bioinformatics, data management and cloud computing is perceived by the respondents, that represent the Portuguese research community to a considerable extent. Most groups are lacking skills on these topics and many lack access to services in their institutions, which explains why training and user support were the most mentioned needs among surveyees.

Data management is the topic where the gap is wider, as we expected from our experience in the [Ready for BioData Management](#) programme, and it should motivate us to extend the training offers in the programme, as well as to widely promote our newly launched [Project Data Management](#) consulting services. Our [bioinformatics services](#), which feature the [GTPB](#) and [crash courses](#) besides analysis and tools development, are also well set to meet the needs of the community. In both cases, however, it is clear we must contemplate a shift in our training offers towards the online modality in order to cater to the preferences of the community and reach a broader audience.

With respect to cloud computing, we are not meeting all the needs of the community, as training, which is currently not provided by our [computing services](#), was the most mentioned need on this topic as well. Maybe other research infrastructures can also contribute to fill this gap.

The perspective of researchers participating in this survey with regard to data sharing and integration is encouraging, and suggests that BioData.pt can play a role in the facilitation of data integration in public repositories through training, user support, and the development of the repositories themselves. Nevertheless, past experience has been that the initial willingness of people for data sharing is seldom substantiated when the time comes to put it in practice.