DTL Focus Meeting

Creating a bioinformatics and data stewardship centre

Summary of discussion

Author: Daphne van Beek.

Ingredients for a successful bioinformatics facility/service centre

The presentations in the meeting gave a broad overview of the different obstacles that someone starting a bioinformatics facility/server might face. Different motivations, strategies, and implementations were shown. A question to the audience was: ‘What are the ingredients that are valuable for implementation?’

- Nobody wants to reinvent the wheel. It seems that a lot of people are willing to share their wheel (which is definitely a positive observation), but there is still little effort to actually develop the wheel together. To efficiently make use of time and money, it is important to find the balance: not duplicating things that are already invented and starting from scratch yourself.
- The importance of bioinformatics should become clearer to the community. Centrally offered services are too expensive for the researchers, as they have so far received many of these services free of charge. Besides the obstacle of charging for services, there is often not enough central money from the institute for dedicated people in or infrastructure for bioinformatics.
- It is important to bill users for bioinformatics services. Make the real costs visible to let people get used to the charge. It is not a problem to give a discount of 80%!
- You can trade scientific recognition for a discount: if people do not want to pay the full price for an interesting project, they can make people from the service a co-author on their paper.
- We anticipate 'cloud coins', 5% of European Commission budget reserved for data services, that can only be used with service providers. This could make sure the money arrives at the right place and services will not be overlooked.
- If we accept that 5% of research budgets is to be used for treatment of data, it is not unreasonable if 5% of the total number of people involved should be in data management/stewardship positions. To convince board members of this fact: visualise the risks if you do not have sufficient capacity or put emphasis on early adoption (in NL, in EU). Remember that money can also be found in the middle layers of your organisation.
- Make sure researchers see the added value of an expert looking in on their project.

Covering the costs for services

The value of data stewardship cannot yet be measured in euros. A data business model is needed to visualise the effects of data stewardship on the future of research. Datasets should be ‘valued’, so influence of the set can be measured. Implementation is hard because most researchers are not yet making the data available.
Is there a general recommendation to the NFU DRE committee? Maybe set up a repository of all data stewardship plans for all UMCs/institutes?

Data quality

During the day, questions often resulted in a discussion about data quality: how can we make sure the quality is sufficiently high and how can we measure it? How can we prevent fake data or poor science finding its way into our repositories?

Focus is on the reusability of *data*, but it would also be preferable to aim for the reusability of *tools*. A lot of tools developed are not used after their development is finished. Is this something to take into account in the FAIR data policy? You are advised to assign a DOI to your software, making it citable and reusable.

Visibility of services offered

The visibility of bioinformatics as a service, as presented by different parties, is poor. All institutes need to be aware that their services are currently not findable by a simple Google search. Websites are often not up-to-date. Sometimes a help-desk exists within the institute, but not all researchers know where to find it or the help desk is not informed about the bioinformatics services.

An important conclusion is that every data service provider in the institute should know the others, and you should refer people to other help desks if you cannot be of service to them.

The FAIR data management tool that is currently being developed in ELIXIR may be customized for local use. This means that instructions can be given to the researchers on where to find services/help if they need it.

Scientific recognition

When is authorship or acknowledgement preferred? What should you consider? If there is creative input into the project from the service centre, scientific co-authorship is aimed for. If the main contribution is running a standard pipeline, one would like an acknowledgement for the work done. When a lot of time is invested in a project, this might be different. The financial gain of the researchers is: you don’t have to invest time to learn bioinformatics yourself; you get experts with broad experience working for you.

Who needs to be FAIR-savvy?

Do we need to educate everyone on the FAIR data principles or just some experts? A large investment is needed, in both time and money.

- Appreciation of FAIR data should be instilled in *all* researchers. Teaching about the general FAIR principles and the main goal might be useful.
- Eventually, we will have automated tools that can make your data quite FAIR for you. Interaction with FAIR data will always be there, so some knowledge should be available in the research field.
- Using workflow management and properly managing legacy data (for example from a PhD student that has left) can increase the efficiency of research.
- Adding FAIR metadata automatically is currently not possible. A specialist is thus needed. Planning pays, because adding all metadata at the end of the project is always harder, as there is a risk things are not recorded correctly and are missed.
- Make sure the researchers see the added value of FAIR data: it makes the data citable and could result in more efficient use of research budgets, by avoiding unnecessarily generating data that already exists elsewhere.
- FAIR should be considered as a ladder (i.e., there are different degrees of FAIRness). No data is completely FAIR or unFAIR. Technology currently under development is a great influencer in the ‘height’ on the ladder that we can reach.