Data and Services Discovery projects - Institutional Role in a Data Commons

Title

Australian Neuroscience Microscopy Data Sharing Platform

Approach

Our goal is to develop Australian Neuroscience Microscopy Data Sharing (ANMDS), a national open-source microscopy data repository. To achieve this long-term goal, we have conducted the following activities:

1. We have formed an ANMDS organising committee which has a total of 22 members. The members include university professors, research group leaders, microscopy specialists, data infrastructure and supercomputing specialists among others and were chosen carefully so that each participating institution in our proposal had a representative in the committee. In addition we have included a postdoctoral researcher as a representative of early career researchers and a PhD student as a representative of graduate students. The committee members have been in regular contact over the duration of the project and have been involved in the preparation and promotion of a national workshop that will be held in Canberra on the 9th of October 2019.

2. We have organized a Data Sharing: Neuroscience, Microscopy and Experiments Symposium. This symposium is a collaborative project between the ANMDS committee and the Australian Characterisation Informatics Committee and the Australian Brain Data Commons Working Group. We have organised morning plenary sessions in which most of the invited international speakers will be discussing different aspects of establishing and running public data-sharing platforms and protocols. The Australian Neuroscience Microscopy Data Sharing, Standards and Best-Practice Workshop will be held in the afternoon on the same day. The workshop is designed to tackle specific problems that may arise with image data curation, accessibility and reusability as well as to shed light on the interplay between standardisation of microscopy data collections and crafting new experimental methods so that all forms of scientific breakthroughs can be facilitated. In the second part of the workshop, in the open discussion forum, we will bring together international and domestic experts to outline the national strategy to integrate current systems into a unified data sharing platform.

More information about the symposium together with the registration link can be found below:
https://bit.ly/2lm7gdW

3. In the run-up to the symposium we have conducted a survey whose purpose was to identify practical challenges and perceived obstacles that researchers have towards data sharing. The survey was sent to all participating institutions and has provided a wealth of information on how researchers feel about data sharing and on ways to overcome difficulties.
The survey can be found at the following link:
https://forms.gle/7kci5fkL3UC9ocRY7

4. We have gathered information on current hardware and software implementations targeted towards data sharing in research institutions across Australia. This information will be essential in outlining the strategy for building a national data sharing platform.

FAIR

The only data sets created in this project are the results of the survey which is still ongoing. We plan to promote the survey at the Data Sharing: Neuroscience, Microscopy and Experiments Symposium, at the Data Summit in Brisbane on the 21st of October, in November at Jason Swedlow's presentation at UQ and at two scientific meetings that will be held between the 1st and the 5th of December. One of the aims of this survey is to bring awareness to the concept of Open Science which is why it is crucial to give time for it to reach as many Australian Scientists as possible. We plan to close the survey by the 10th of December 2019. The results will then be analyzed and made publicly available according to FAIR principles.

Collaboration

Our project initially involved collaboration between nine Australian Universities and six institutes. Together we have reached out to international experts who will share their expertise on how to best implement a national data-sharing platform and provide methods of value measurement to estimate the significance and impact of this platform. In addition, we have expanded the collaboration to include the Australian Characterisation Informatics Committee, the Australian Brain Data Commons Working Group, Microscopy Australia, MASSIVE, the National Imaging Facility and the National Collaborative Research Infrastructure Strategy network. The output is a collaborative symposium focused on data sharing and characterisation. In addition, we have strengthened existing ties and have started to form new ties with international experts. Professor Jason Swedlow will be visiting the University of Queensland in November to give a presentation and discuss OMERO and the Image Data Resource in the context of a national data sharing platform and public image data work. David Orloff, the director of Cell Image Library, has openly expressed interest in long-term collaboration.

Sustainability

At the Australian Neuroscience Microscopy Data Sharing, Standards and Best-Practice Workshop we will develop a national strategy for the development of an ANMDS platform and create a two-year plan of action. This plan will include setting up the infrastructure and recruiting scientists to deposit their data so that modifications can be made and issues addressed based on

early-user feedback. The ANMDS committee will oversee all activities and ensure compatibility with already existing international platforms.

In terms of infrastructure and platforms that manage data, significant work has been done to provide platforms that are FAIR in their intent and could house data supporting FAIR standards. There is a growing realisation that FAIR data sharing platforms are lifecycle and process driven (data is not born FAIR). As a result, UQ has created the RDM (Research Data Manager) [https://rdm.uq.edu.au] and a fabric, Metropolitan Data Caching Infrastructure (MeDiCI), which are not only supporting FAIR principles and workflow approaches, but they are also nationally and internationally extensible. Whilst the platform, fabric and overarching data governance model are very important, the integration of repositories atop these fabrics is critical for a data sharing platform. Without all these key elements being in place infrastructurally, sharing with appropriate governance and assurance of standards would be a best effort only.

To disseminate the outcomes of this project we have reached out to Light Microscopy Australia who have promoted the microscopy data-sharing initiative and collaborative symposium to their members and this has significantly increased participant numbers. In addition, Dr Durisic will announce the initiative to create an ANMDS platform in December at the Australian Biophysical Society, the Australian Electrophysiological Society and the Australasian Neuroscience Society meetings during her oral presentations and present at the Data and Services Summit in Brisbane on 21 October 2019.

Learnings

Our findings show that many of the researchers in Australia are not aware of data sharing capabilities in their institutions (yes 44%, 56% no or maybe) and only 41% of the researchers have shared their data in some way. Most of those who did share their data have used platforms available in their institutions (80%). The other 20% have used internet solutions such as Gdrive and DropBox. The speed of upload and whether the platform was user-friendly was of concern for all researchers who shared their data. The responses of those who did not share their data showed that they were new to research or have never been thinking about data sharing. Majority (65.4%) said that they would share their published microscopy data if we had an Australian Microscopy Data Sharing Platform. However, most of the researchers said that they would not share their data if the data would be available for everyone to download and use. Only 27% of researchers in Australia have reused data generated by other scientists and if they did, the data was created by someone they knew and trusted. Trust that the quality of other peoples' data would be at a satisfactory level and a lack of relevant data in public repositories were the main obstacles in reusing data. Surprisingly, most researchers in Australia were not familiar with the FAIR principles.

So what can be done to increase data sharing? Our findings show that what is most important is to educate Australian scientists and provide competent support on data management and data sharing. This is most important for young researchers who are at the beginning of their careers and are in the process of learning to generate large quantities of data. It is also important to create

fast and straightforward routes to data sharing as many scientists, especially those with advanced career stages, see a lack of time as the main obstacles to making their data public.

Impact

This project has mainly involved the Australian Optical Microscopy community but has been naturally extended to researchers who generate data in other fields. Several research groups at the University of Queensland have already changed their approach to data management and adopted a project-based framework through the use of the UQ RDM and MeDiCI. This has increased not only their ability to manage data more efficiently but has also enhanced their collaboration with national and international partners, significantly increasing their overall research productivity. Since the RDM is relatively new at UQ, the involvement of the researchers has also helped the UQ Research and Computing Team to fine-tune data access and movement and the locality of data to computational resources. This concerted effort between researchers and infrastructure specialists has enabled the creation of an environment truly capable of supporting FAIR principles in scientific data management and has already led to the publication of data in conjunction with a journal article. This practice is expected to grow over time and is further facilitated by Nature Scientific Data, a new Nature Journal dedicated to publishing data, whose Senior Editor, Associate Professor Kaylene Simpson, will attend the Data Sharing: Neuroscience, Microscopy and Experiments Symposium. The symposium was designed to fully elaborate on all aspects of FAIR principles for scientific data management and stewardship. We believe that raising awareness of these principles will transform the way scientific communities operate, which will in turn increase reproducibility in science and boost the value of Australian scientific data.

Report prepared by: Nela Durisic and Rumelo Amor, Queensland Brain Institute, The University of Queensland.
Date: 08/10/2019