

## To B2SHARE or not to B2SHARE?

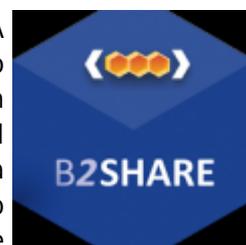
Last month EUDAT's roving reporter caught up with Carl Johan Håkansson, the current product manager for EUDAT's B2SHARE service, who is based at the PDC Center for High Performance Computing at the KTH Royal Institute of Technology in snowy Stockholm.

### ***Good morning, Carl Johan. Would you please tell us what B2SHARE is?***

Yes, sure. B2SHARE is a service for European academic and industrial researchers and research communities, as well as citizen scientists, to store, share and publish long tail research data.

### ***Before you go any further Carl Johan, would you clarify what you mean by “publishing” data and also explain what a “service” is in this context?***

Ok, there are two main ways that researchers may need to share data with other people. A researcher might be collaborating with one or more colleagues and have some data to “share” (in the sense of all using or changing that data) during their ongoing research process, but, until the research has reached a certain point, those researchers would probably not want to make the data available to anyone else. The second kind of data sharing happens when researchers are ready to make their data publicly available to everyone, that is, they want to “publish” the data. In the latter case, **B2SHARE** is the service they should choose as the data stored in it can be searched for and viewed by anyone with access to the web.



In the former case, **B2DROP** (which is another of EUDAT's services) is the appropriate choice – it is a bit like Dropbox in that researchers can store data in B2DROP and then allow specific people to access that data. Think of it as though B2DROP is the place to store data while you are working on it and possibly changing it, and then, when you are ready for everyone and anyone to look at and use the data, you move it to the B2SHARE service.

Now, you had asked what I meant by a “service”. In computer terms, the distinguishing feature of a service is that you can use it without having to buy or install any special software – basically all you need is a web browser that lets you access the web.

***Thanks, so when you say B2SHARE is a service for storing, sharing and publishing research data, the good thing is that people don't need to buy any extra software, they can just go ahead and store their research data directly from their favourite web browser? And anyone who wants to search for data can just do that straight from their browser too?***

Yes, exactly. Researchers can simply start up Chrome, Firefox, Internet Explorer or whichever web browser they prefer and go directly to <https://b2share.eudat.eu> and then either start searching the published data that is already there, or login and deposit some new data. Of course, if researchers have not used the B2SHARE service before then they will need to choose the option to sign up for an account before using the service. However, the process of getting a confirmation of registration should only take a few minutes, so people can start almost straight away.

I should also add that we have another way that researchers can store data in B2SHARE or get copies of data stored there. It quite often happens that people need to store or use a large number of data files. When this happens, it is too time consuming to sit and deal with each file by hand via the web. So EUDAT has also made it possible for people to access B2SHARE using things known as application programming interfaces, or APIs. These can be used, for example, to automatically transfer a large volume of data files between a research community's website and B2SHARE. However, using B2SHARE in this way is something that will usually be handled by a data

manager, rather than researchers – so don't get too worried about APIs if you are not familiar with them!



**Thanks for that clarification, Carl Johan. So European researchers from academia or from industry or even ordinary people who are doing some research can register and put their data into storage in the B2SHARE service or search there for some interesting data they'd like to use. That brings us to the question of the kind of data that can be stored in B2SHARE. You mentioned something about long tail data, but what do you really mean by that? (I assume EUDAT hasn't set up a whole service for storing information just about kangaroos, peacocks, crocodiles and other long-tailed creatures...)**

Well, to understand long tail data, I'm afraid you need to forget kangaroos, and instead think about the different kinds of research data that exist these days. Some people work with instruments that produce lots of measurements – for example, think about all the astronomical data being recorded by telescopes or the climate and seismological data from recording stations. Other researchers run simulations that generate large amounts of data, such as when modelling a boat in motion to design a better hull shape, or modelling how parts of the human brain work. As you can imagine, the large amounts of data that are used in these kinds of research are too big to be stored on a USB stick or personal hard drive, so they need to be archived at a data centre with facilities for storing vast quantities of data. However there is also a lot of research that works with sets of data that are small enough to be stored on a laptop or in a university mainframe computer – for example, someone might have medical survey results that are kept in Excel files or digital photos documenting observations of rare plants or birds in the wild.

It is this latter kind of data, which comes in relatively small quantities, that is known as “long tail data”. The reason that it is called long tail data is that if we drew a diagram of the size of all the sets of research data, and arranged them in descending order by the size of each set of data, we would have a relatively small number of large sets of data on one side of the diagram – the climate and astronomical measurements – and a very large number of sets of small data on the other side – the bird photos and medical survey results and so forth. You can imagine that the myriad small quantities of data form a shape like a long tail on the diagram. The use of the term “long tail” actually comes from the retail sector where it is used to refer to the large number of products that only sell in small quantities (like Lamborghinis, Lotuses and Aston Martins), in comparison to the smaller number of items that sell in large quantities (such as Hyundais and Volkswagens).

Now the thing is that all of this research data is important – whether it comes in large or small sets – but the problem is that the small sets of information stored on a camera or laptop or a departmental computer can be lost, for example, when the person who created the data moves on to another university, or when the laptop is replaced. Another problem is that it is usually hard for other researchers to find data that is stored on an individual's personal laptop or stored in a local departmental computer system at another university!

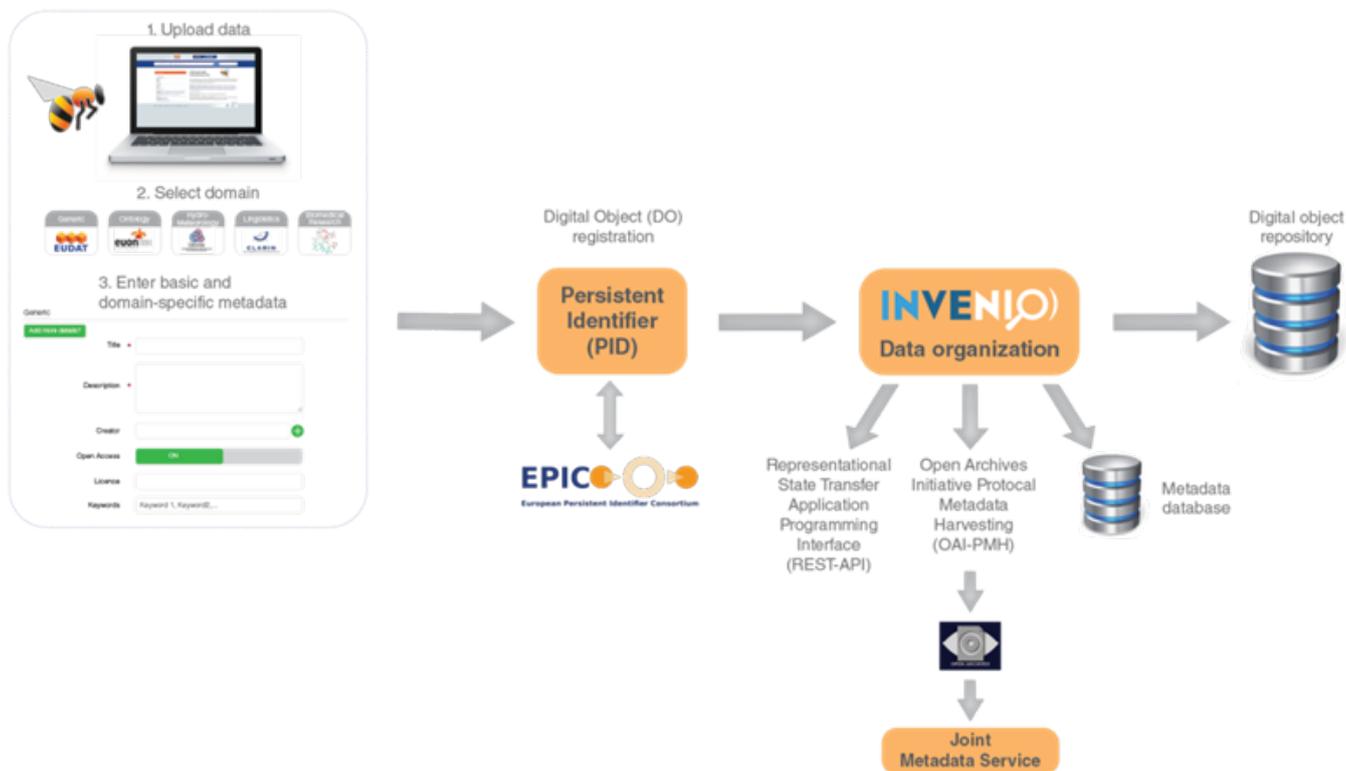
So the idea of B2SHARE is to make it possible for researchers to store these relatively small sets of data somewhere where other researchers will be able to find and use the data, and in such a way that the data will continue to be available in the long-term.

***Well, we definitely do need to stop data like that evaporating into thin air, Carl Johan, and that certainly makes B2SHARE useful for lots of researchers! But now I'm left wondering how B2SHARE actually works. It seems like a very big task to provide such a large amount of storage and make it accessible to researchers all over Europe. Where do you actually store all this data? How do you make it all happen?***

Yes, it is a big task, however it isn't actually EUDAT that provides all the computer systems for storing the data.

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.

What we do is to develop the software (that is, the actual program code) that makes the B2SHARE service work, and then we provide that code to research data centres so they can make the service, and hence storage, available. At present, B2SHARE is just being run at the CSC-IT Center for Science in Finland, however we are working with research communities and institutions around Europe to get more and more research data centres involved with providing the B2SHARE service. So EUDAT is serving a dual role - firstly, as a software developer/provider, and secondly as a facilitator to establish a network of data centres in Europe offering storage for long tail research data.



**Thanks for the explanation, Carl Johan. One final question: if EUDAT is not actually providing the computer systems for storing the B2SHARE data, how reliable will this service be, and how can researchers who use B2SHARE be sure the data will continue to be available in the future?**

I'm glad you mentioned that as reliability is a very, very important feature - the whole point of the B2SHARE service is to make sure that long tail research data is not lost! So, do you remember that I mentioned we are working on having a large number of European data centres making the B2SHARE service available to researchers? The idea here is that the data centres and research institutions that participate in the B2SHARE service will make a commitment to keep all the data that is deposited in their storage via B2SHARE available for a particular length of time. The agreement will also include a commitment to keeping B2SHARE available for a further two years if a centre decides to shut down their B2SHARE service. This will allow time for another data centre to take over responsibility for that lot of B2SHARE data and move the data to their own site, so it will continue to be available.

**Thanks very much, Carl Johan. We are out of time for today, but I'll be getting back to you later on to ask more about the practicalities of using the B2SHARE service.**



I look forward to it. Meanwhile anyone who would like to make a start on sharing some of their long tail research data is welcome to sign up now at <https://b2share.eudat.eu>. If anyone needs help, there are also links there to documentation about B2SHARE and information on how to contact support.

[Read more](#)