

Storing irreplaceable data and making it perpetually available for future researchers

Andreas Taeger, Senckenberg Institute

Key Points

- **Raw data sharing**

Sharing raw data, and not being limited to the amount of data you can share, only makes research better.

- **Perpetual data storage**

Storing data that is difficult or impossible to replace offers a sustainable, perpetual solution.

About Andreas

I am a biologist working in entomology within taxonomy. We are trying to recognise the diversity of Sawflies which are not flies, but rather Hymenoptera, like bees and so on. My work is to try to find the correct names of the species.

In this special group, there are about 8,000 species which are known around the world. As part of this, I try to, more or less, clean up the names and recognise undiscovered species. The information on type and specimen is housed everywhere in the world - in big museums, usually - and the material may be 250 years old or older. Of course, new species we discover today have new specimens. But our science is going back to Carolus Linnaeus in 1758 - the first man who used Latin names for animals. We are still working with this original material - everybody who has been working since this time is using such material. Therefore, we have a lot of old material, but also, of course, new.

Researching and storing irreplaceable data

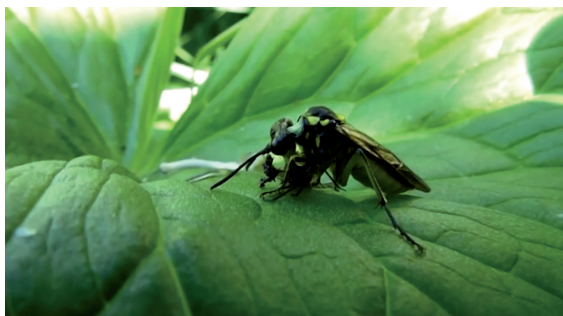
This material is difficult to examine. You have to go to the museums or borrow the material which is shipped over here from the USA or Great Britain or Germany. And so, it's always a rather expensive thing to get this material for examination simply to check that its name is still valid. There have been many, many insects and animals with more than one name in the past. We compare the original types and check that they're exactly the same. If they aren't, we have to fuse the names into one species. We want this information to be available for the future. There are so many insects and it's very difficult to repeat the same work again and again.

Also, materials can become damaged during travel and information can be lost. Content on figshare is published - there is a lot of information on there about the structures which will not get lost. This information is unique - there is only one specimen in the world. If this is lost, it's very complicated to replace it. It's practically impossible.

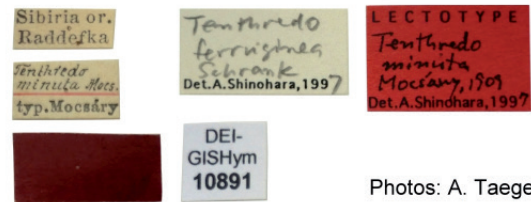
“Content on figshare is published - there is a lot of information on there about the structures (it’s not complete, but it is at least a very important part of the research) which will not get lost. These so-called types, they are unique - there is only one specimen in the world. If this is lost, it’s very complicated to replace it. It’s practically impossible.”

Using a DOI to link to research

On the other hand, we are publishing the research on paper. In the past, it was almost impossible to include colour plates because it’s very expensive and there were very few copies available. Today, we have lots of good technical equipment to take pictures or scanning microscopes. We can take very high resolution pictures. This is what I do as part of my research - examine specimens and take pictures. There is no way to publish this data in a journal, at least not without loss. If you have a picture that is 10,000 pixels, you cannot print it on a small page - you simply lose the information. Therefore, I take all my original pictures and put them in figshare for publication. I don’t publish the images in the journal, but rather give a link to the DOI number of the



The sawfly *Tenthredo mariae* Aguado, 2014, is feeding a fly



Photos: A. Taeger

Tenthredo minuta Mocsáry, 1909, lectotype

image. For me, this is the most important thing - that it has a DOI number. This is a safe way to keep these pictures. We are a big institute but I will eventually leave once I’ve finished my work. I’m not sure if the original photos will survive after 5 or 10 years if they’re on my old computer or the network. I think it’s important that this data is publically available, which is why I’m using figshare. I can upload all the data I would never publish because I know it will be available in 100 years.

Sharing more data means a better research output

The process of my research has changed in that I’m using more information in my papers and publications as I’m including more or less infinite figshare references. For example, I have a paper where I included a reference to at least 50 plates, but in reality, I only published 10 plates in the journal. This was not possible in the past. In this case, the output of my work is getting better.

Get in touch:

figshare.com

info@figshare.com