

Digitising Natural History Collections

CSIRO National Science Experiment: Technology in the CSIRO National Collections

What is happening?

CSIRO maintains the National Research Collections Australia, which contain over 15 million specimens of plants, insects, fish, wildlife, algae and tree seeds.

The National Research Collections Australia are currently being digitised. Digitising is the process of creating a digital copy of a physical object, such as a physical specimen, a photo or a video, so that it can be stored by a computer. Computers store data as a series of binary digits, made up entirely of 1s and 0s. This storage method makes them very good at doing maths and storing text but gets far more complicated when we're looking at storing images or representations of physical objects.

As well as digitising specimens by taking high resolution photos of them, we digitise the data associated with each specimen. This data is written on the specimen's label. It includes information such as the date it was collected, the species (identified by an expert!), the latitude and longitude where it was collected, and even information like the kinds of plants growing nearby.



Machine Learning and Artificial Intelligence can be used to get information about specimens from images. © CSIRO

How we store collections

Physical specimen collections are very important to science. They allow us to understand the biodiversity of an area, and how plants and animals have changed over time.

The data attached to a specimen which include where it was collected can help us understand how the places different plants and animals live has changed over time.

Many specimens are delicate and need to be carefully maintained. For example, the Australian National Herbarium stores over a million plant specimens. Most of these plants are stored on sheets of archival paper, but some specimens where the 3D structure of the specimen is important to researchers, such as orchid flowers, are stored in vials of ethanol, which is a liquid used to preserve specimens by stopping the growth of bacteria and preventing decay. To store a plant on archival paper, first they are pressed in a plant press and dried in a special oven, then frozen to kill any insects or insect eggs that are on them, then attached to the paper either by using tape or by sewing them on.

The specimens are then stored inside folders of similar specimens, flat on shelves, inside the Herbarium building. The collections continue to grow over time as researchers add more specimens, which include both brand new species and new specimens that reveal the diversity of known species.

Why we digitise

We are digitising our collections to make them available online, for free, to people all over the world. They can be used by scientists for all different kinds of research, by students who want to know more about particular plants and animals, and even by community groups who want to restore ecosystems.

No specimen can be replaced if they are lost – each specimen is a unique record of biodiversity. Digitising specimens provides some additional security in case the original specimen is lost, damaged or destroyed. But there is so much more a physical specimen can tell us about biodiversity!

Having a digital copy of a specimen makes it easier to share our collections with researchers in Australia and around the globe. Having quick access to information that can help with bushfire recovery and biosecurity will mean that we can quickly respond to threats to our natural environment and work quickly to protect our delicate ecosystems. Digital copies of specimens can be shared with researchers instantly across the globe.

The collections act a bit like a library, loaning specimens to researchers at other collections around the world and borrowing specimens from other collections as needed. Being able to share specimens digitally means we can loan them instantaneously and to more than one person at a time, which cannot be done with physical specimens. We can also share the specimens without risking their safety in transit.

Digitisation also adds another layer of protection for the specimen. If the digital versions of the specimen can be used for scientific research instead of the physical versions, it means that the delicate samples are moved less often, meaning that there is less risk of damage.

Having specimens stored digitally also allows us to search through the collections and find the specimens we're looking for more quickly, as well as enabling artificial intelligence and machine learning tools to quickly analyse the files that we've stored and look for common features and patterns in the millions of available samples that humans may not notice.



Plant specimens from the cinnamon group loaded on the Picturae system © CSIRO



A researcher taking images of an insect using a high-definition camera setup © CSIRO

How we store collections

The digitisation of the Australian National Herbarium was done using a specially designed conveyor belt system which was made by a company called Picturae.

The system works much like an airport security scanner, with a conveyor belt passing under a camera that takes a high-resolution photo. This system enables archivists to work faster by quickly moving specimens through the imaging system and linking them to details that are stored on a barcode which links the physical specimen to its description. Once specimens are on the conveyor belt, they pass under a high-resolution camera, which takes a very detailed photo. This photo is checked by a researcher for errors before the specimen is packed away.

Animal specimens can often be tougher to capture, depending on what the specimen is, and the way it is stored. The types of specimens stored in the collections are from a wide range, which include eggs, insects, algae, tree seeds, sound recordings, birds and bone samples. The Australian National Insect Collection contains more than 300,000 microscope slides, which are comparatively easy to digitise, since the specimens have been flattened, and only need to be photographed from one side.

Other samples, including pinned insects often require multiple high-resolution images to capture what it looks like from multiple angles and make sure that all key features are captured. Taking these photos quickly requires a specially designed camera rig, which enables the specimen to be captured from multiple sides without touching it.

Challenges

There are many challenges to digitising collections. One of the major challenges is adding data from handwritten or typed notes, which have often become faded and discoloured over time.

Some of the slides in the National Insect Collection up to 100 years old! Using tools such as optical character recognition (OCR) is a way to speed up the process of digitising these written records. OCR is a process that uses a computer algorithm to analyse a picture, looking for patterns that indicate written text. These tools can often use artificial intelligence and machine learning to improve their accuracy.

Another enormous challenge is the amount of time required to digitise all of the specimens. CSIRO's natural history collections contain more than 15 million specimens. Some of the specimens need restoration as part of their digitisation process, which is time consuming. Even if no additional work was needed and each of the specimens in the collection only took one minute to digitise, it would mean working 24 hours a day, 7 days a week for more than 28 years to digitise every single specimen.



Collections trivia!

- A few of the plant specimens in the Australian National Herbarium were collected by Joseph Banks during Cook's voyage to Australia in 1770.
- The Australian National Algae Culture Collection and the Australian Tree Seed Centre both store living specimens!