

Progress report on C3S Data Rescue for ACRE Antarctica
 (based at NIWA, New Zealand)
 Andrew Lorrey, NIWA Auckland

Workshop for C3S Data Rescue held at NIWA Auckland

The C3S Data Rescue workshop for capacity building was held in Auckland during December 2017. There were two full days of training on tools and research presentations, as well as a half-day field trip held in conjunction with the 10th ACRE Meeting. The full programme can be found here: <https://www.niwa.co.nz/node/111840>

Platform for digitising old weather observations

The citizen science website “Southern Weather Discovery” will go live in late September 2018. This new website is hosted by Zooniverse and has essentially utilized the format employed recently by Ed Hawkins “Weather Rescue” project. It will help us to capture tens of thousands of additional observations through the portal in quintuplicate, which is required for quality control. We have arranged the site so that several foci for data keying can be promoted (station data, ship log books, sledging missions, etc). The first beta test of this platform was essentially completed in late August 2018 and there are some instructions for participants that are being changed to make the data entry easier.



- Support from C3S allowed us to “copy” Ed Hawkins “Weather Rescue” effort.

WORDS FROM THE RESEARCHER



“We have about 150,000 images of ship logs from archives in the UK and Scandanavia. Each image has several days’ observations and each day has multiple observations so we have millions of observations to key over the course of the project.”

- Citizen scientists volunteer to go on a “mission” - categorized by data type (eg. whaling, exploration, sledging)
- We will also use “Old Weather” for non-tabulated observations in diary format
- Data uploads will target events we would like to examine that are poor in 20CR



Weather Data detection and AI

NIWA is currently investing in use of Ephesoft Transact software, which is used to detect the presence of data in old meteorological registers. This will allow us to derive a time series that shows temporal depth of observations from recovered ship log books, lending to improved targeting for data keying. Ephesoft is a partner of Microsoft, who has also taken an interest in our project. They will be using several examples of our rescued data to build a character library that can be called on using AI to potentially undertake automated data transcription using OCR/ICR. This type of effort will be tested using a small data set, and also pushed to the Southern Weather Discovery citizen science website as a way to determine AI accuracy for data recovery.

