



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Kuwait data Management Training

**Date Rescue and
role of WMO RCCs**

1-10-2025

Marlies van der Schee



Data rescue problem

Extent of problem:

- › Archives in uncontrolled climatic environments
- › Risk of loss by extreme events
- › Even today NMHSs write observations in notebooks without a clear push to have this digitized





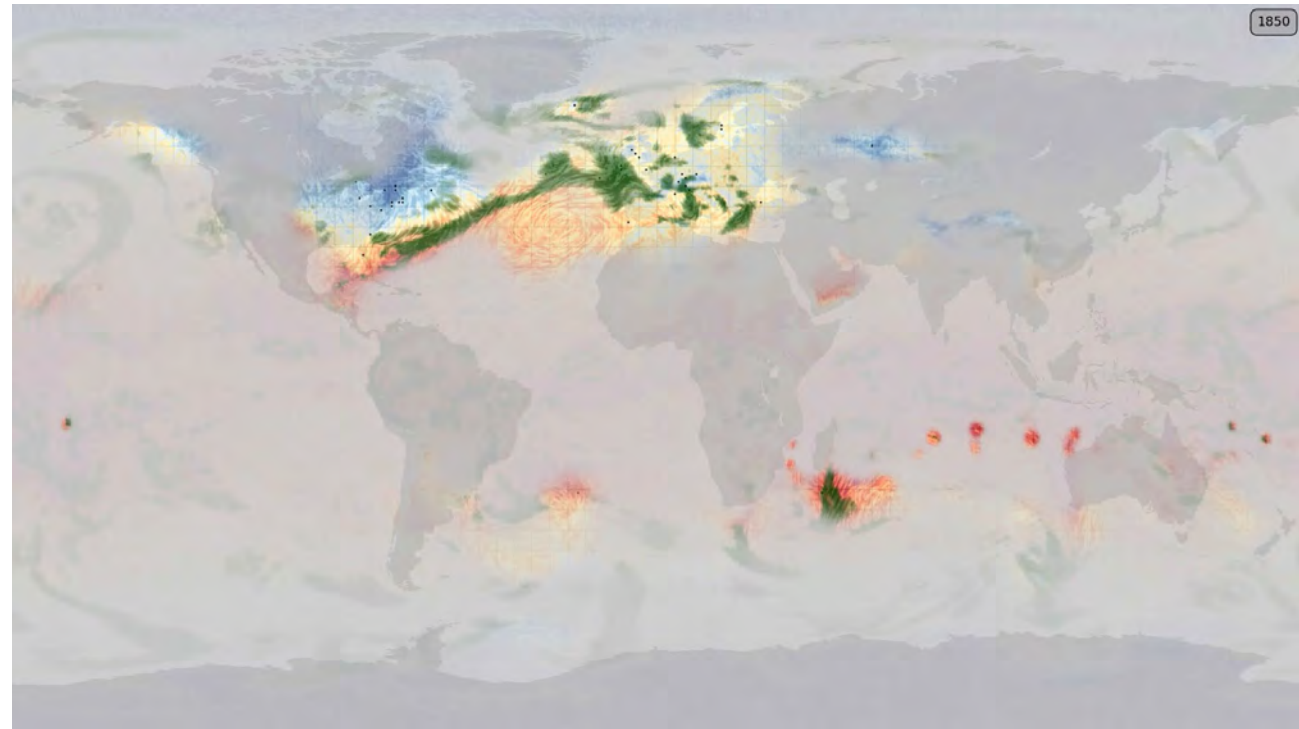
Things that can go wrong after successful DARE project

- Physical back-ups go lost
- Co-workers from retirees/former personnel do not know what has been done
- Rescue is done twice from copies of original data from e.g. Bulletins colonial history

Importance Data Rescue

Historical data is used for climate analyses:

- › Climate models such as ERA5, E-OBS and 20CRv3
- › Climate change attribution
- › Including more data increases our understanding on climate and extreme weather events

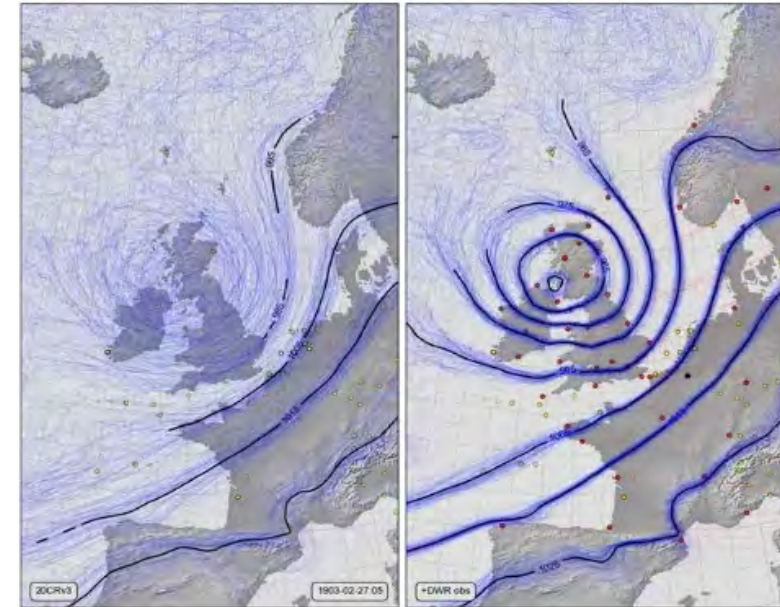


20CRv3: grey fog masks indicate high uncertainty (Temperature)

Importance Data Rescue

Historical data is used for climate analyses:

- › Climate models such as ERA5, E-OBS and 20CRv3
- › Climate change attribution
- › Including more data increases our understanding on climate and extreme weather events



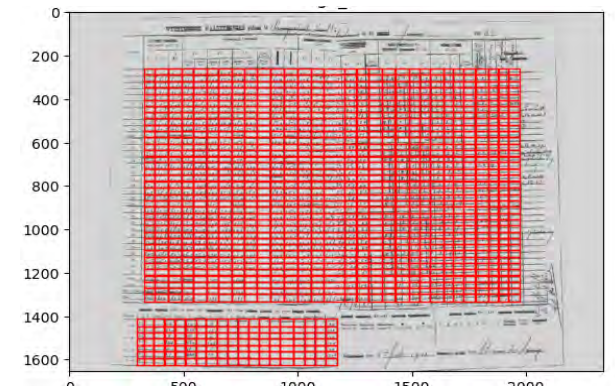
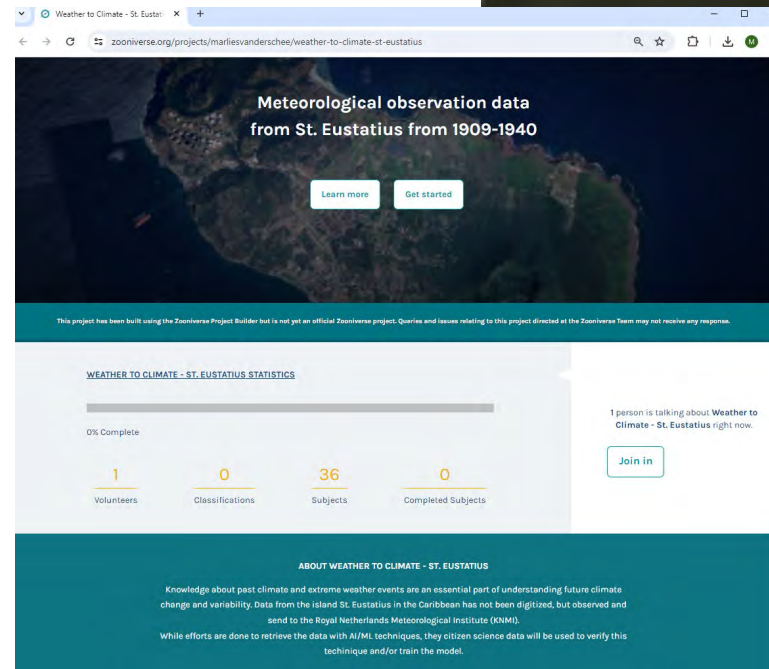
Inlcuding 44 stations (red dots) increases the understanding of the pressure contours of the Ulysses storm in 1903, UK.



Options for digitization

1. Manual annotation
2. Citizen science (Zooniverse)

Future: Optical character recognition an AI/ML/DL?





Step 1 – Digitize sheets

RDWD Innovation week set-up:

- › Camera set-up according to WMO data rescue best practices:
 - Use glass plate to avoid folds
 - Name images correctly with software program
 - Store images securely with off-side backup
 - Create image inventory
- › ~72 sheets/hour
- › RAW image files





Step 1 – Digitize sheets

RDWD Innovation week set-up:

- › Camera set-up according to WMO data rescue best practices:
 - Use glass plate to avoid folds
 - Name images correctly with software program
 - Store images securely with off-side backup
 - Create image inventory
- › ~72 sheets/hour
- › RAW image files





Max	Min
25,6	21,2
25,4	21



1. Manual annotation

1. Contracting local people

2. Students:

- Part of Curriculum:
 - Hands-on exercise
 - Learn useful transferrable skill
 - Feeling of a worthwhile contribution to global project
 - Example of classroom materials: <https://datarescue.climate.copernicus.eu/classroom-materials-climate-data-rescue-africa-project-clidar-project> (Noone et al., 2004)
- Useful paid work and possibility for them to learn from work at Meteo Service:



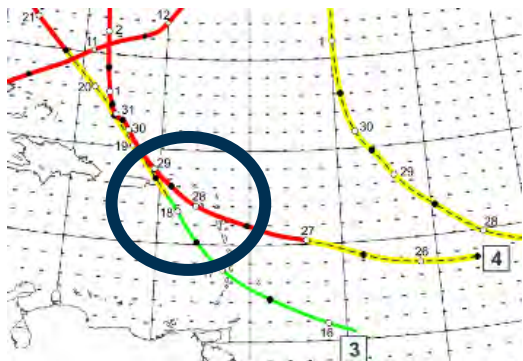
1. Manual annotation

1. Contracting people from in other countries via online platforms:
 - <https://www.upwork.com/> provides a platform in people completed many other jobs
 - Experience for 4 sheets from St. Eustatius (example picture included on sheet):
 - Job was picked up in 5 hours
 - work completed in 48 hours (over the weekend)
 - \$10 dollars were paid – however, bargaining was done rather softly

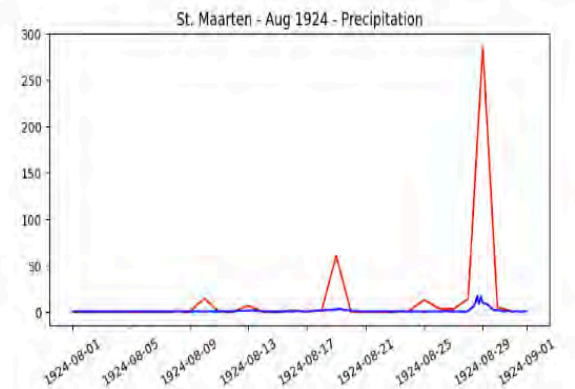
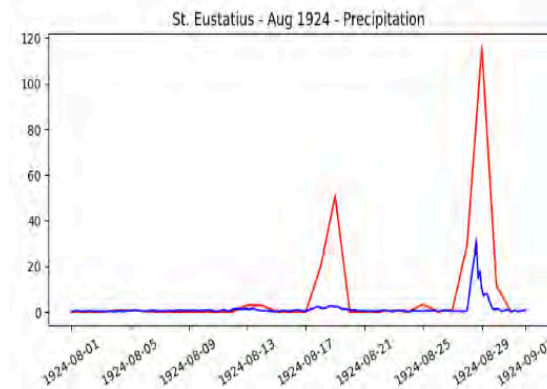
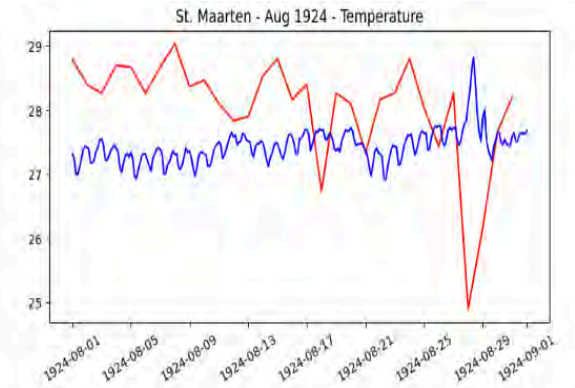
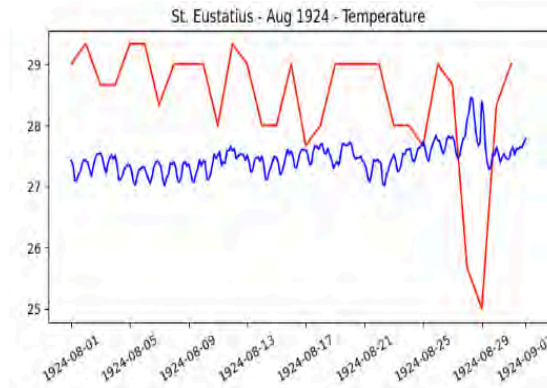


Example: 20CRv3 - observations

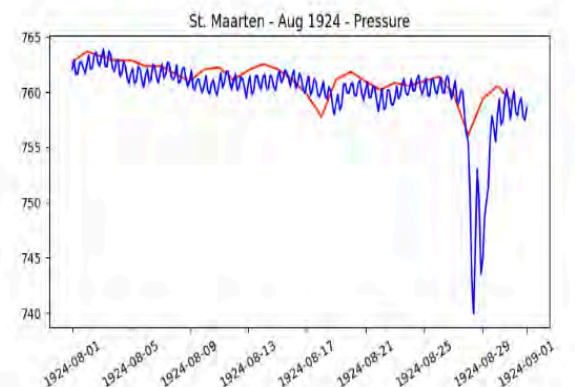
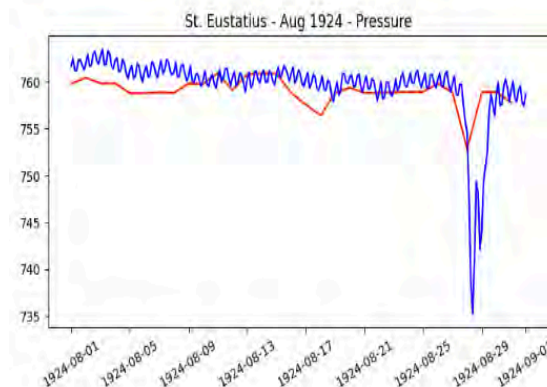
- › Identified hurricane: August 1924
- › 20CR grids are too large to represent the smaller BES-islands
- › Difference with 20CRv3:
 - Temperature is overestimated
 - Precipitation underestimated
 - Pressure underestimated in observations



Orkaan.
De orkaan die verleden week hier geseind was en in westelijke richting tusschen de eilanden Dominica en Antigua gepasseerd is, heeft ook aan onze Bovenwindsche eilanden een bezoek gebracht en volgens bericht van den Gezaghebber ontvangen er nog al aardig huisgehouden.



Archive
Reanalysis



2 Citizen science (Zooniverse)

1. Workflow established for 1911 precipitation
2. The output can be used for climate science, AI/ML

<https://www.zooniverse.org/projects/marliesvanderschee/weather-to-climate-st-eustatius-caribbean>



A screenshot of a web browser showing the Zooniverse project page for 'Weather to Climate - St. Eustatius'. The page features a dark header with a satellite image of St. Eustatius and the title 'Meteorological observation data from St. Eustatius from 1909-1940'. Below the header, there are two buttons: 'Learn more' and 'Get started'. A teal banner below the header contains a disclaimer. The main content area has a light blue background and includes a progress bar showing '0% Complete'. Below the progress bar, there are four statistics: '1 Volunteers', '0 Classifications', '36 Subjects', and '0 Completed Subjects'. To the right, it says '1 person is talking about Weather to Climate - St. Eustatius right now.' and has a 'Join in' button. At the bottom, there is a teal section titled 'ABOUT WEATHER TO CLIMATE - ST. EUSTATIUS' with text explaining the project's goal to digitize historical meteorological data from St. Eustatius and use AI/ML techniques to verify the data.

2 Citizen science (Zooniverse)



1. Create your own workflows

PROJECT #24038

View project

Project details

About

Collaborators

Field guide

Tutorial

Media

Visibility

Talk

Data Exports

Workflows

Subject Sets

Translations

NEED SOME HELP?

Road a tutorial

Ask for help on talk

Glossary

OTHER ACTIONS

Delete this project

Precipitation 1911 in THE FARM / PLEASURE #26806

A workflow is the sequence of tasks that you're asking volunteers to perform. For example, you might want to ask volunteers to answer questions about your images, or to mark features in your images, or both.

WORKFLOW TITLE

Precipitation 1911 in THE FARM / PLEASURE

If you let your volunteers choose which workflow to attempt, this text will appear as an option on the project front page.

TASKS

How many mM 'Neerslag' are recorded on 'Datum' 1 in the FARM	T0
How many mM 'Neerslag' are recorded on 'Datum' 2 in the FARM	T1
How many mM 'Neerslag' are recorded on 'Datum' 3 in the FARM	T2
Which 'maand' is recorded?	T3
How many mM 'Neerslag' are recorded on 'Datum' 4 in the FARM	T4
How many mM 'Neerslag' are recorded on 'Datum' 5 in the FARM	T5
How many mM 'Neerslag' are recorded on 'Datum' 6 in the FARM	T6
How many mM 'Neerslag' are recorded on 'Datum' 7 in the FARM	T7
How many mM 'Neerslag' are recorded on 'Datum' 8 in the FARM	T8
How many mM 'Neerslag' are recorded on 'Datum' 9 in the FARM	T9
How many mM 'Neerslag' are recorded on 'Datum' 10 in the FARM	T10
What has been the 'Som' of 'Neerslag' from day 1-10?	T12
Does the header of the last column state 'The Farm' or 'Pleasure'?	T13
Which year is recorded - please, indicate the full year (19XX) <i>(first)</i>	T15

Add a task

FIRST TASK

Which year is recorded - please, indicate the full year (19XX)

MAIN TEXT

Which year is recorded - please, indicate the full year (19XX)

Describe the task, or ask the question, in a way that is clear to a non-expert. Note: this is a label for the text input; markdown can be used only to add images (with alt text), bold and italic text.

HELP TEXT

You can find the year at the top right of the page.

![Example Alt Text]

Add text and images for a help window.

☒ Required

METADATA TAGS

Volunteers can attach the following tags to highlighted portions of their transcription.

☐ Deletion

☐ Insertion

☒ Unclear

NEXT TASK

Which 'maand' is recorded?

Delete this task

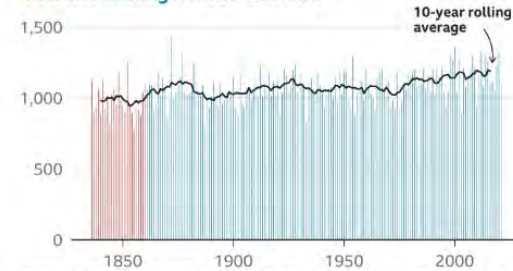


Citizen Science – Best example

- › UK MetOffice/University of Reading
- › Covid lockdown
- › 16,000 volunteers
- › 65,000 scanned pieces
- › 5.4 million rainfall observations

UK citizens extend rainfall record

Annual millimeters of rainfall between 1836 and 2021, **new** and **existing** records combined



Source: UK Met Office/Uni of Reading/Rainfall Rescue



Scientists have been amazed at the public's response to help digitise the UK's old rainfall records.

Handwritten numbers on documents dating back 200 years are being transferred to a spreadsheet format so that computers can analyse past weather patterns.

The volunteers blitzed their way through rain gauge data from the 1950s, 40s and 30s in just four days.

Project leader Prof Ed Hawkins had suggested the work might be a good way for people to use self-isolation time.

"It's been incredible. I thought we might get this far after three or four weeks, not three or four days," he told BBC News.



AI/ML problem

First reaction: Easy problem to solve, right?

- › Workshop AI for Data Rescue, March 2024
- › E-AI WG10
- › Specific publications on AI and DARE without good solution

Literature review on Copernicus website

AI/ML and Data Rescue

Artificial Intelligence has attracted a lot of interest in the media after the launch of ChatGPT. This triggered meteorologists to look for practical solutions to solve the problem of handwritten or printed meteorological journals with observations. What makes Data Rescue complicated is that unlike simple optical character recognition (OCR) of printed prose, we are also interested in the structural layout of the text in tables and digitizing the handwritten text within them.

Currently, The Swedish Meteorological and Hydrological Institute (SMHI) launched a project named [Dawsonia](#), acronym for Digitize hAndWritten observationS in weather jourNals, with the ambition to optimize and train a sufficiently accurate machine learning model which can handle different forms of tabular data, convert handwritten-text and produce machine-readable files.

KNMI published illustrative notebook in [GitHub repository](#), including a manual table overlay, and adding sharper contrasts between cells as a pre-processing step for Dawsonia.

[Lorrey et al. \(2022\)](#) used the Zooniverse platform to obtain a transcribed dataset to test Microsoft Read API. The Microsoft Read API validation grand strike rate was $69\% \pm 15\%$ ($n = 920$). They suggest that collaborating with industry to enhance optical character recognition (OCR) capability has the benefit of accelerating data rescue progress that can rapidly augment scientific data repositories. A high quality video on their project is available at [Vimeo](#).

[Singh and Middleton \(2024\)](#) recently published a scientific article on the issues of Tabular Structure Recognition.

Philip Brohan, working at the UK MetOffice, conducted a study using [Gemini 2.0](#) Flash Experimental multimodal large language model to extract monthly precipitation data from the UK Ten Year Rainfall Reports. Near-perfect accuracy was achieved. He concludes that Gemini is extremely capable for this use case, and impressively easy to use. It will take some more work to get good results for more complicated cases with for example variable missing data annotated with a short line. However, it looks as if it is ready to go as-is for simpler cases with no missing data. Read more on his [webpage](#) on this experiment.

[Zhang et al. \(2024\)](#) shows that their preliminary workflow offers a foundation for future attempts to automate the transcription of historical records, which are available on [GitHub](#).

Workshop: AI for Data Rescue

26th March 2024
10:00 - 15:00 UK Time

Online



Fringe 2024

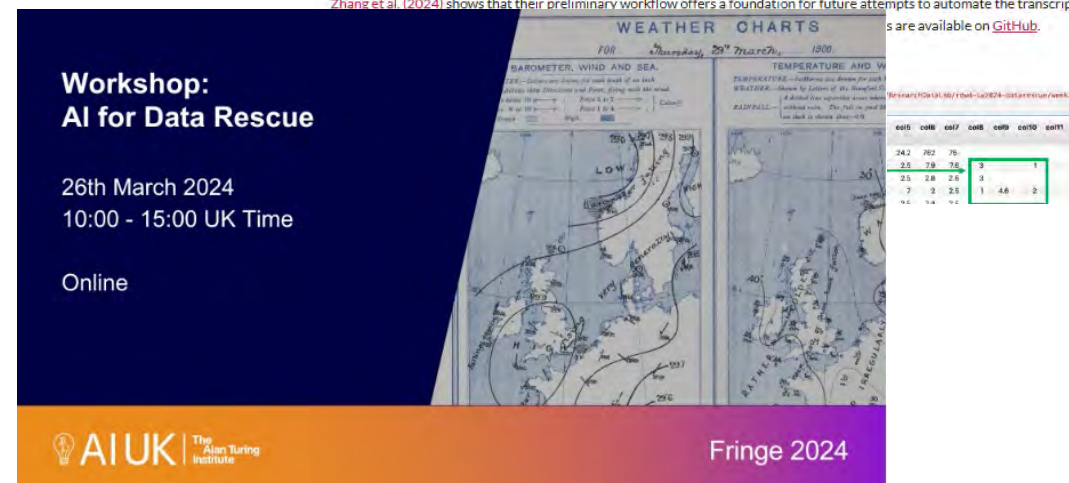
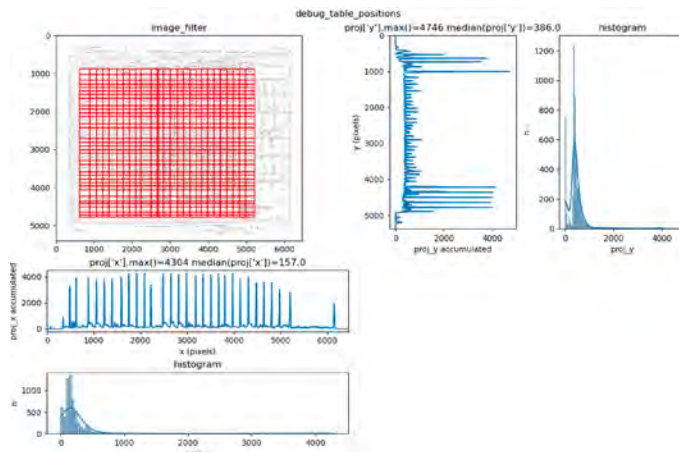
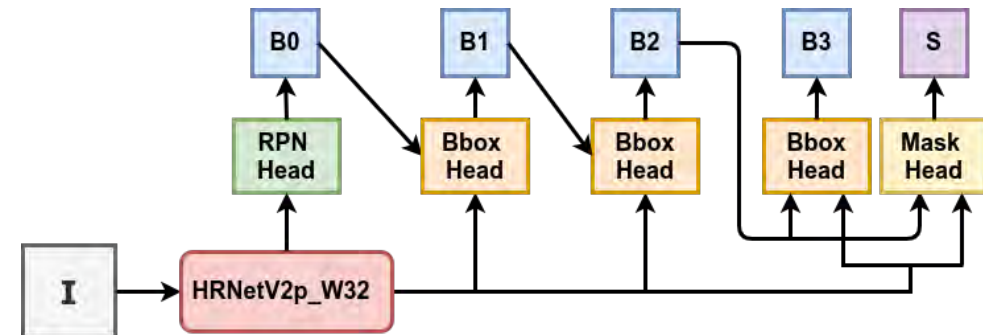


Table Structure Recognition (TSR) followed by Optical Character Recognition (OCR)

- › Use two deep learning models:
 - One to identify table structure
 - Other to read entries from table cells



- › CascadeTabNet
- › Semi-supervised learning



CascadeTabNet - Prasad et al. (2020)



- Table recognition

DATUM	LICHTSTERKING van half 6 op 10 1/2			TEMPERATUUR °C								DAMPSTERKING mm			BETREKKELEZE VOCHTHEID %			WINDRICHTING EN KRACHT in gsm 1/2			BEWOLKING 0-100			REKEN- SLAG in 200 afmeten in 100				
	7.30	13.30	19.30	7.30		13.30		19.30	Damp 7.30-13.30 op 19.30	Maxi- mm	Mini- mm	7.30	13.30	19.30	Damp 7.30-13.30 op 19.30	7.30	13.30	19.30	7.30	13.30	19.30	Damp 7.30-13.30 op 19.30						
				droge bol	natte bol	droge bol	natte bol																droge bol		natte bol			
1	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
2	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
3	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
4	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
5	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
6	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
7	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
8	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
9	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
10	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
11	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
12	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
13	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
14	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
15	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
16	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
17	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
18	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
19	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
20	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
21	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
22	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
23	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
24	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
25	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
26	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
27	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
28	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
29	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
30	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
31	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
32	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
33	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
34	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
35	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
36	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
37	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
38	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
39	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
40	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
41	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
42	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
43	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
44	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
45	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
46	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
47	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
48	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2.6	2.8	3.0
49	20.55	20.44	20.21	25.6	26.7	28.2	16.6	23.6				70.2	23.2	22.8	21.5	26.2	23.6	23.6	6.3	5.5	5.9	24.8	1.1	2.2	2.4	2		



MeteoSaver v1.0

- › 74% matches the manually transcribed record

Recommendation:

- › Enhance the robustness of the table and cell detection
- › Improve transcription accuracy
- › Expand QA/QC check

derrickmuheki Added a manual for MeteoSaver v1.0 a095266 · 6 months ago 120 Commits		
📁 .vscode	Updated modules	last year
📁 OCR_HTR_models	Final changes for draft manuscript	10 months ago
📁 data	Final changes for draft manuscript	10 months ago
📁 docs	Add files via upload	10 months ago
📁 manual_and_minimal_working_example	Added a manual for MeteoSaver v1.0	6 months ago
📁 results	Validation plots - changes	10 months ago
📁 src	Final changes to script before manuscript submission	10 months ago
📄 .gitignore	package structure	last year
📄 Dockerfile	Dockerfile working	last year
📄 LICENSE	LICENSE	last year
📄 README.md	README.md	10 months ago
📄 configuration.ini	Final changes for draft manuscript	10 months ago
📄 environment.yml	Add files via upload	10 months ago
📄 job_script.sh	Job_script.sh added for users using hpc infrastructure	10 months ago
📄 setup.py	package structure	last year
📖 README 📄 AGPL-3.0 license		



AI extraction - Gemini

- › Input: response schema
- › Works perfectly...
- › Unless entries are missing
- › Commercial software with estimated \$0.25 per page

A.M. Form 1003, Air Ministry, Meteorological Office. Station No. 1678

Station Number: 4007
Observer: Theodor Carlsson & Co. Ltd.

Year: 1947
Station Number: 1678
Location: BADWORTHY COTTAGE, S. BRENT
County: DEVON
Height above sea-level: 550
Gauge diameter: 5
Gauge height: 1 ft 3 in

REGISTER OF RAINFALL IN 1947

Station No. 1678
Location: BADWORTHY COTTAGE, S. BRENT
County: DEVON
Height above sea-level: 550
Gauge diameter: 5
Gauge height: 1 ft 3 in

THE MORNING MEASUREMENT SHOULD BE ENTERED TO THE PREVIOUS DAY

Please read the Notes and Instructions Overleaf

Date	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Total
1	0.54	1.19	-	0.05	0.14	-	-	0.08	-	0.07	-	-	1
2	0.31	0.42	-	0.01	0.45	-	-	0.06	-	0.63	0.08	-	3
3	0.40	0.67	-	0.56	0.02	0.42	-	1.59	-	-	x	20	3
4	-	0.11	0.03	0.10	0.11	0.39	-	39	-	65	0.02	-	5
5	1.03	-	0.79	1.60	0.08	1.10	-	-	-	-	-	32	4
6	0.43	-	-	0.34	0.12	0.22	-	22	41	-	-	13	5
7	0.04	0.29	0.15	0.32	0.01	0.02	-	64	11	-	02	61	5
8	0.62	-	-	0.15	-	0.58	-	58	-	-	75	09	5
9	0.06	0.17	1.23	-	0.14	0.52	-	52	-	36	87	-	9
10	1.24	-	1.08	-	-	0.37	-	37	-	-	26	-	9
11	0.49	0.01	0.28	-	-	-	-	-	41	-	x	-	5
12	0.33	-	2.28	-	0.23	0.11	-	11	1.33	-	09	-	5
13	1.05	-	0.35	-	-	-	-	03	-	-	-	-	7
14	0.11	-	0.17	0.05	0.52	-	-	-	-	-	72	06	7
15	-	1.74	-	0.06	-	-	-	-	-	-	03	-	7
16	0.15	0.12	-	-	-	-	-	-	-	-	-	-	5
17	0.05	0.00	-	0.74	0.05	-	-	-	-	-	43	-	5
18	-	0.40	-	-	0.04	-	-	-	-	-	64	-	18
19	-	0.36	0.12	-	-	-	-	-	-	-	18	-	18
20	-	0.65	1.01	-	-	1.50	-	1.50	-	-	28	-	18
21	-	0.17	0.16	-	-	-	-	75	-	-	35	65	18
22	-	0.55	0.18	-	-	-	-	-	28	-	1.29	48	1
23	-	0.39	0.91	-	0.02	-	-	-	1.41	-	04	05	1
24	-	0.05	0.01	0.52	-	-	-	-	-	-	12	02	4
25	-	1.00	-	0.05	-	0.03	-	03	-	-	13	28	4
26	-	-	-	-	-	-	-	-	-	-	-	1.20	4
27	-	-	0.45	-	-	0.51	-	-	-	-	-	1.19	4
28	0.02	0.22	-	-	0.56	-	03	-	-	-	-	08	17
29	-	-	0.11	-	-	-	-	-	-	-	05	11	17
30	0.63	0.32	0.47	0.37	-	0.08	-	-	-	-	04	04	17
31	-	-	0.04	-	-	-	-	-	-	-	-	94	17
Total	7.48	3.44	15.48	5.89	3.22	5.62	5.24	2.13	3.73	2.17	7.05	7.42	

Philip Brohan (Met Office)
https://brohan.org/AI_daily_precip



Conclusion AI/ML

- › Methods have not been proven to be good enough to be used with the high accuracy that our community is looking for.
- › Advance in this field are rapid.
- › Can we use a combination of AI with human interference?
 - Manually check the values
 - Include manually table detection



Climate
Change

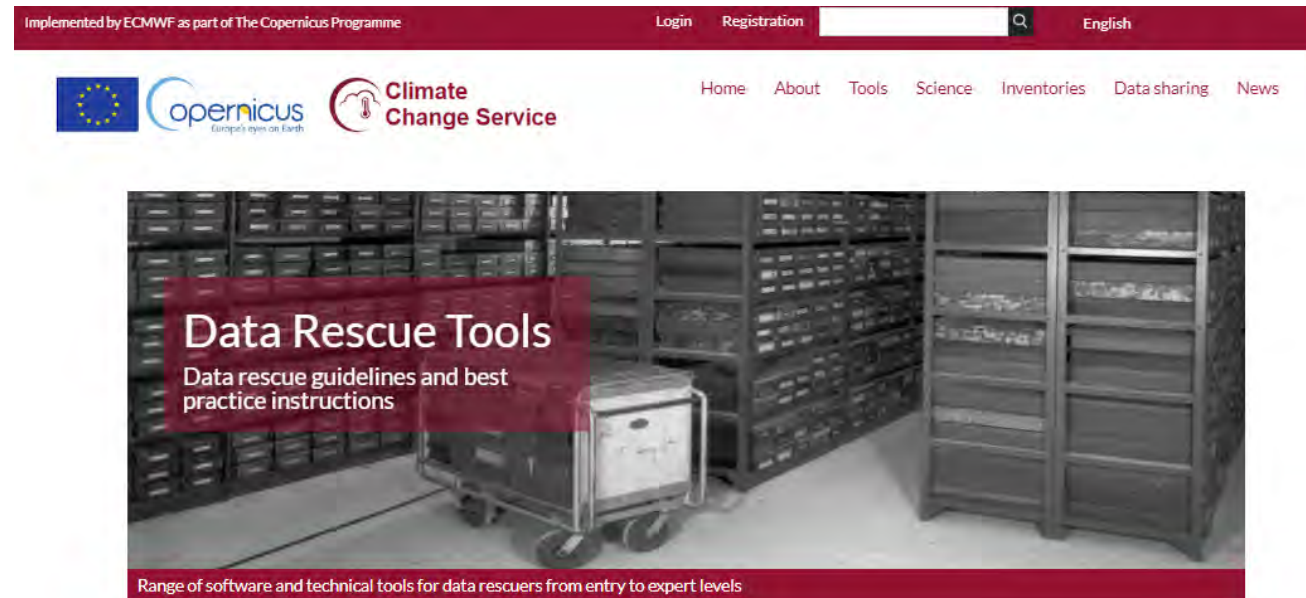
Data Rescue Portal



Mission :

- Share information as guidance
- Single entry point on current data rescue projects
- Enabling collaboration among organizations, development agencies, donors, scientists NGOs, etc.
- Accelerate data rescue in support of climate assessment and adaptation

Homepage DARE portal



Old WMO I-DARE portal

The data rescue portal is designed to facilitate and coordinate the rescue of weather and climate data from around the world. The service runs an online user-based system that provides access to information on past, current and planned data rescue projects, as well as tools and guidance to facilitate each stage of the data rescue process.

Mission of the Data Rescue Portal:

- a collaborative framework on sharing information, best practices, know-how, guidance, metadata on data rescue projects and activities worldwide.
- provides a single entry point for accessing information on the status of climate data being digitized or in need of recovery and digitization.
- enables collaboration among organizations, development agencies, donors, scientists, NGOs, citizens, to work on the data recovery and digitization of climate heritage which is at risk of loss forever.
- through above and capacity building responds to the need for accelerating data rescue in support of climate assessment and adaptation, risk management and disaster risk reduction.

The practical information, data rescue projects and metadata inventories originate from initiative of both the [World Meteorological Organization](#) (WMO) and [Copernicus Climate](#)

News

Upcoming presentation DARE Portal at EMS, Barcelona

2024-08-26

The new data rescue portal will be presented at the EMS Annual Meeting in the session on [Climate Monitoring, data rescue, management, quality and homogenization](#) on Friday afternoon the 6th September. Read the [abstract](#) on the website of the EMS.

Marlies van der Schee will present on the changes from the last portal and she will...



Climate
Change

Share information as guidance

- References to success stories, including citizen science projects
- References to current scientific data rescue projects and advances
- Software tools
- Links to important guidelines on data rescue, including the update DARE guidelines
- News items on DARE events or project highlights

Implemented by ECMWF as part of The Copernicus Programme

Login

Registration



English



Copernicus
Europe's eyes on Earth



Climate
Change Service

Home About Tools Science Inventories Data sharing News

Success Stories

ACRE

The international [Atmospheric Circulation Reconstructions over the Earth \(ACRE\)](#) initiative both undertakes and facilitates the recovery of historical instrumental surface terrestrial and marine global weather observations to underpin 3D weather reconstructions (reanalyses) spanning the last 200-250 years for climate applications and impacts needs worldwide.

- [ACRE and ACRE-facilitated 20CR Overview](#)

- [ACRE structure and map](#)

- [Allan, R. et al., 2016: Toward integrated historical climate research: the example of Atmospheric Circulation Reconstructions over the Earth, WIREs Clim Change](#)

For the latest work in ACRE, please read [this](#) news item.

HISKLIM

The KNMI-program '[HISKLIM](#)' (HIStorical CLIMate) has the objective to make historical climate data physically accessible for both land and sea, from sources in the Dutch language, with the highest possible temporal resolution and of the best possible quality.

[Historical Canadian Climate Data: Volunteer Data Rescue Project](#)

Over the past ten years, with the help of organizations such as the Network in Canadian History and Environment (NICHE), the Meteorological Service of Canada (MSC), McGill University, and the McCord Museum archives, early weather observations have been copied and digitally imaged. Many of the records are from Eastern Canada, where most of the early settlers lived.

IEDRO

The mission of IEDRO, the [International Environmental Data Rescue Organization](#), is to locate, rescue (image), digitize, archive and share historic weather data across the globe, in areas of need, for the purpose of supporting applications in agriculture, infrastructure planning, disease prevention and climate change.

- [Current, future, and completed projects](#)

ISTI

The [International Surface Temperature Initiative](#) has over the past 5 years worked with numerous international partners to collate existing and rescued holdings into its first databank release which has served to increase the data completeness and station count throughout the long-term record. Although the databank first release consists solely of monthly mean temperatures submission of daily and sub-daily data and non-temperature elements are encouraged. Such data is shared with other relevant holdings such as the International Surface Pressure Databank, Integrated Synoptic Database and Global Historical Climatology Network-Daily. Rescued data under various historical and ongoing DARE projects has been incorporated or is in the process of being incorporated. The Initiative also aims to promote further data rescue of land meteorological data and is willing to partner in such activities or provide letters of support to proposals. It has a particular interest in development and implementation of crowdsourcing approaches to the problem which are likely the most cost effective approach. ISTI can be reached via general.enquiries@surfacetemperatures.org and databank submissions of rescued data can be made by contacting data.submission@surfacetemperatures.org to arrange submission.



PROGRAMME OF
THE EUROPEAN UNION



IMPLEMENTED BY





Climate
Change

Share information as guidance

- References to success stories, including citizen science projects
- References to current scientific data rescue projects and advances
- Software tools
- Links to important guidelines on data rescue, including the update DARE guidelines
- News items on DARE events or project highlights

Implemented by ECMWF as part of The Copernicus Programme

Login Registration English

Home About Tools Science Inventories Data sharing News

Software Tools

On this page there are recommendations for an array of software tools to facilitate all the steps in the data rescue process. There is also information about the tools, their application, and ancillary files that support their use.

- [Organise/edit images](#)
- [Optical Character Recognition \(OCR\)](#)
- [Digitisation of strip charts](#)
- [Speech recognition](#)
- [Version control](#)
- [Formatting \(land stations\)](#)
- [Formatting \(marine data\)](#)
- [Data quality control](#)
- [Metadata quality control](#)
- [AI/ML for data rescue](#)

   WORLD METEOROLOGICAL ORGANIZATION

This Portal has been set up and will be further developed in close collaboration with WMO.



PROGRAMME OF
THE EUROPEAN UNION





Climate
Change

Share information as guidance

- References to success stories, including citizen science projects
- References to current scientific data rescue projects and advances
- Software tools
- Links to important guidelines on data rescue, including the update DARE guidelines
- News items on DARE events or project highlights

News



[Consultancy post for DARE for Djibouti](#)

2024-09-03

The WMO has opened up a consultancy post for data rescue work for Djibouti. The expert will assess the current status of data recovery activities in Djibouti, both within the Agence Nationale de la Météorologie (ANM) and at the Centre d'Études et de Recherche de Djibouti (CERD), and develop an implementation plan to advance data recovery in accordance with the guidance provided in the Good Practice Guidelines for Data Recovery, 2024 Edition (WMO-No. 1182, WMO

[Upcoming presentation DARE Portal at EMS, Barcelona](#)

2024-08-26

The new data rescue portal will be presented at the EMS Annual Meeting in the session on [Climate Monitoring: data rescue, management, quality and homogenization](#) on Friday afternoon the 6th September. Read the [abstract](#) on the website of the... [more](#)

[International ACRE Drought and Flood Conference, 23-26 September](#)

2024-08-07

ACRE is organizing a 3-day [ACRE meeting](#) on the island of Corfu in Greece to discuss droughts and floods. These enhance stresses and induces wide ranging impacts across both natural ecosystems and human societies, with losses of life, habitat, drinking water and food production. However, the great bulk of drought and flood studies have... [more](#)

News items at DARE portal

edition)



PROGRAMME OF
THE EUROPEAN UNION





Climate
Change

DARE projects

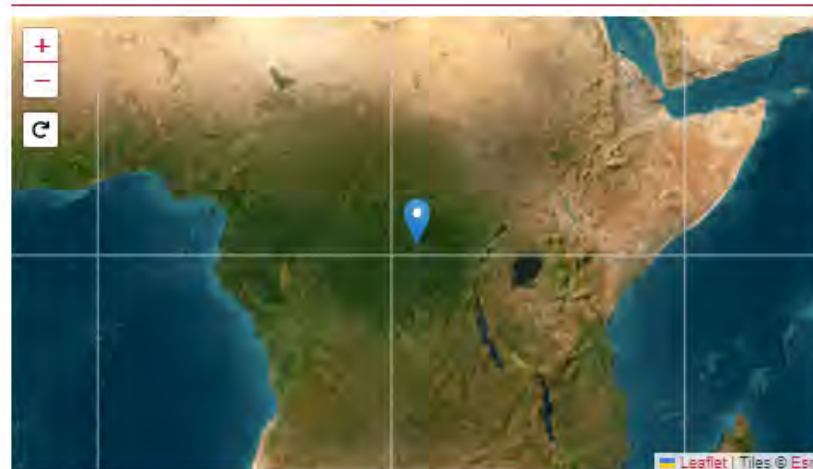
Share your project, ensures rescuing data only once!

- Let the world know on your project
- Make sure data is not rescued twice
- Ask for assistance
- Reference for your project and find possible donors
- 150 projects are available in the portal



Hydroclimatic Data Rescue Mission in the Democratic Republic of the Congo

Project information



Description

Our project focuses on the transcription of a long-term dataset of hydroclimatic time series data for the Congo Basin, currently preserved in archives as handwritten records. This dataset includes daily records of precipitation and temperature from the 1960s within the Congo Basin.

At the start of this project, we conducted a comprehensive inventory of all available hard-copy records of precipitation and temperature hosted in the archives of L'Institut National pour l'Étude et la Recherche Agronomiques (INERA) in Yangambi, Democratic Republic of the Congo (DRC). These archives contain post-1960 hydroclimate data recorded at 37 meteorological stations across the DRC. We developed detailed metadata for all the available data.

Following the inventory, we launched a data digitization campaign, scanning over 9,000 hard copies of hydroclimate records from the archives in both Yangambi and

Features

Contact information (email address)

derrick.muheki@vub.be

Rescue Status

75%

Rescue Status Date

Mon, 09/02/2024 - 12:00

Need assistance

No

Country

[Congo](#)

Institute (or other) name

Vrije Universiteit Brussel and L'Institut National pour l'Étude et la Recherche Agronomiques (INERA)

WMO Region

[WMO Region I \(Africa\)](#)

More information

This project is led by the Department of Water and Climate at Vrije Universiteit Brussel and is funded by the *Research Foundation Flanders*, Fonds Wetenschappelijk Onderzoek (FWO) (grant no.



Share via C3S data deposition service

- Will be included to the Re-analyses of ERA7, on which many climate analyses are done

<https://datadeposit.climate.copernicus.eu/home/>

The screenshot shows the header of the Copernicus Climate Change Service website. It includes the 'Climate Change Service' logo with the URL 'climate.copernicus.eu' and the 'Copernicus Europe's eyes on Earth' logo. Navigation links for 'Home' and 'Sign in' are present. The main heading is 'Global Land and Marine Observations Database Upload Service'. Below this, the section 'Upload Service' explains that the database provides integrated global land and marine surface meteorological holdings from multiple sources. It states that the data ranges from sub-daily to monthly in frequency and is global in coverage. A paragraph mentions that the Upload Service allows data providers from anywhere in the world to contribute. A list of four steps is provided for contributing data: 1. Contact us to request an account and/or to find out if your rescued observations are unique (i.e., not already available). 2. Complete a simple form to confirm you have the required minimum metadata needed for our inventory. 3. Complete a metadata form for each Data Collection you wish to contribute. 4. Upload files/directories as required for each Data Collection. A paragraph notes that the Global Land and Marine Database is a collaboration between the C3S and the National Centers for Environmental Information (NCEI, <https://www.ncsl.noaa.gov>), and that the data submitted will be securely held by both the C3S and NCEI (the designated World Data Centre for Meteorology). The section 'Would you like to provide data?' asks if the user wants to provide data and provides instructions for requesting an account or logging in. The 'Methods of uploading data' section lists three methods: A web form, Rsync, and FTP. A red button labeled 'Start uploading' is at the bottom.

Climate Change Service
climate.copernicus.eu

Copernicus
Europe's eyes on Earth

Home Sign in

Global Land and Marine Observations Database Upload Service

Upload Service

The Global Land and Marine Observations Database provides integrated global land and marine surface meteorological holdings from multiple sources. Various meteorological parameters are served together through a single interface in support of the Copernicus Climate Change Service (C3S). The data ranges from sub-daily to monthly in frequency and is global in coverage.

This Upload Service allows data providers from anywhere in the world to contribute to our database.

In order to contribute either land or marine observations you will need to:

1. Contact us to request an account and/or to find out if your rescued observations are unique (i.e., not already available).
2. Complete a simple form to confirm you have the required minimum metadata needed for our inventory
3. Complete a metadata form for each Data Collection you wish to contribute
4. Upload files/directories as required for each Data Collection

The Global Land and Marine Database is a collaboration between the C3S and the National Centers for Environmental Information (NCEI, <https://www.ncsl.noaa.gov>). The data submitted will be securely held by both the C3S and NCEI (the designated World Data Centre for Meteorology).

Would you like to provide data?

If the answer is "yes" then please [request an upload account](#)

If you already have an account then please [login](#) to begin uploading data.

Methods of uploading data

The data can be uploaded by:

- A web form
- Rsync
- FTP

[Start uploading](#)



Role WMO Regional Climate Centers (RCC)

- › *“WMO Regional Climate Centers (RCCs) are centres of excellence that create regional products including long-range forecasts that support regional and national climate activities, and thereby strengthen the capacity of WMO Members in a given region to deliver better climate services to national users.”*





WMO RCC Functions

Mandatory:

- › Operational Activities for Long Range Forecasting (LRF)
- › Operational Activities for Climate Monitoring (CM)
- › Operational Data Services, to support LRF and CM
- › Training in the use of operational RCC products and services

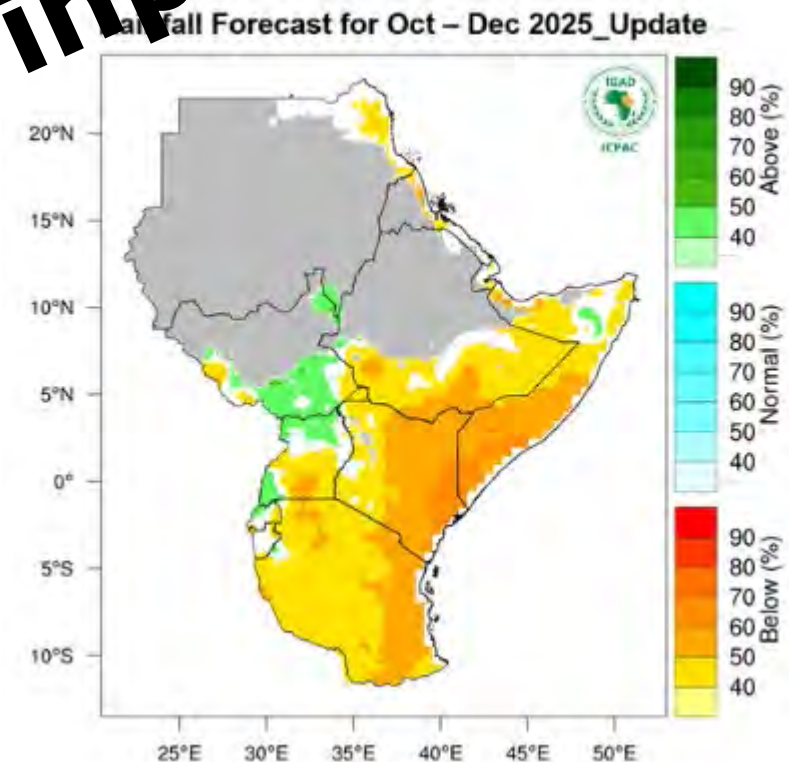
- › Highly recommended:
- › Climate prediction and projection (beyond 2 yrs)
- › Non-operational data services
- › Coordination functions
- › Training and capacity building
- › R&D



Highlight Products: Seasonal forecast

- › Seasonal forecast is one of the most important products used from RCCs
- › Dissemination at Climate Outlook Forum

Such products need input data

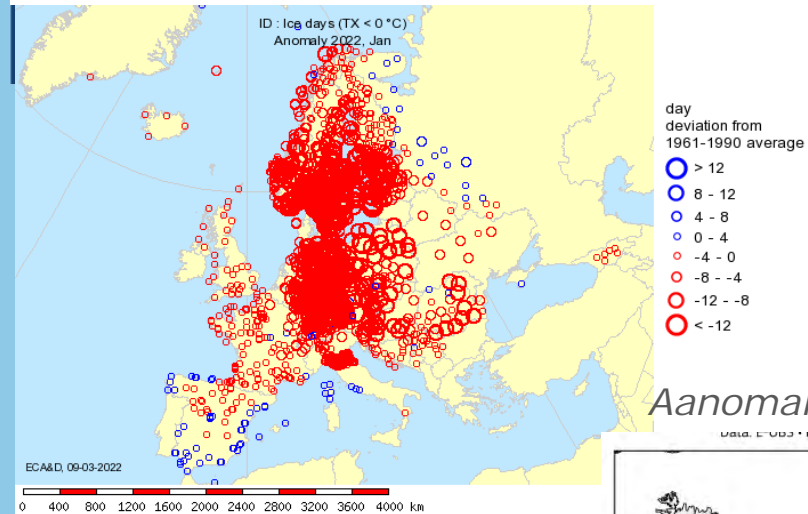


ECA&D

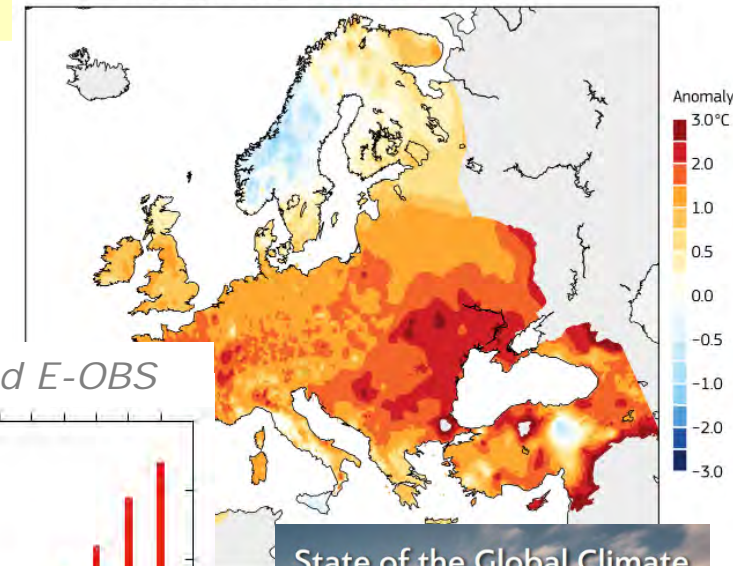
European Climate Assessment & Dataset:

- Initiated in 1998 and maintained by KNMI
- Largest climate daily station dataset from Europe
- WMO - Regional Climate Centre European Climate Data – **obligation of RCC to gather climate data**
- 85 data providers over 65 countries from ~25.000 stations
- European (C3S) and WMO State of the Climate report
- Gridded dataset E-OBS

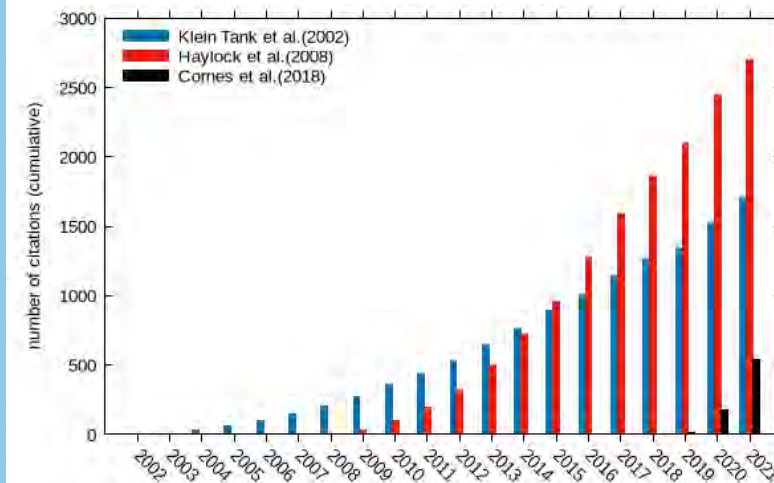
Ice days anomaly, Jan 2022



E-OBS,
Anomalies air temperature 2023



Citations ECA&D and E-OBS



State of the Global Climate
2023



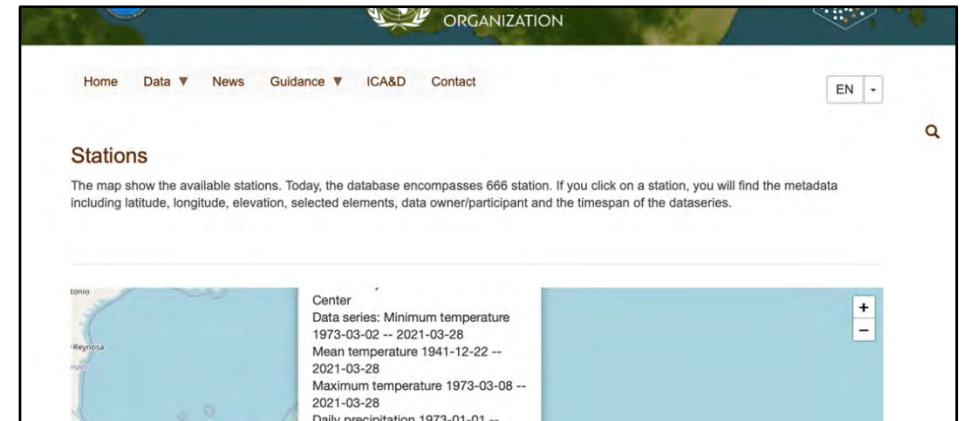


RCC-Node Climate Data services


Features:

- Basic historical climate observations with daily resolution including metadata
- QC procedures and homogeneity check
- Derived indices of climate extremes
- Location map
- Timeseries
- Custom Query
- News


<https://caribbean.icad-wmo.org/>



31 countries selected

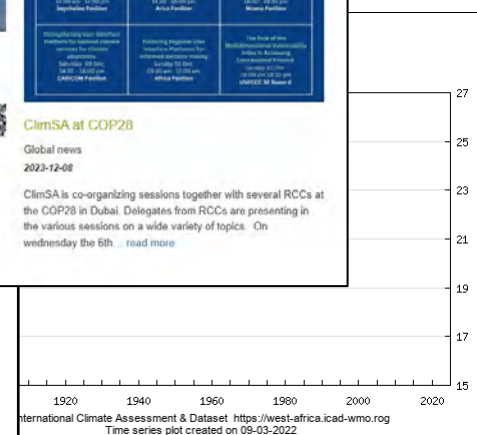
Location 

658 stations selected

Element 

5 elements selected

☐ Additional selection criteria



Features on website



Thank you



Marlies.van.der.schee@knmi.nl