



PROGRAMME OF
THE EUROPEAN UNION



IMPLEMENTED BY



#EUSpace

Facilitating data rescue initiatives

Els van den Besselaar, Marlies van der Schee,
Gerard van der Schrier, Teun Peeters

KNMI, The Netherlands

Introduction – Data Rescue

Extension of problem:

- Numerous archives in uncontrolled climatic environments
- Data is at risk to be damaged by extreme events
- Even today NMHSs write observations in notebooks without a clear push to have this digitized



Current scheme of communicating observations and saving them

Archives of NMHSs in Southeast Asia



Example data sheet from St. Eustatius

Step 1 - Data Rescue Portal

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 currently available in the portal
- <https://datarescue.climate.copernicus.eu/>
(Maintained in C3S with KNMI as WP lead)



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 Kinshasa, DRC. Additionally, we provided training for INERA staff on digitization techniques for archived climate data, furthering our collaboration and capacity-building efforts with INERA.

Currently, we are in the process of transcribing the digitized data (data images) into

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. 11M8825N). We are working in collaboration with the Institut National pour l'Étude et la Recherche Agronomiques (INERA), Isotope Bioscience Laboratory, CAVELab and the Ghent Centre for Digital Humanities at Ghent

Step 1 – Data Rescue Portal

Homepage DARE portal



- Mission :
- Share WMO DARE guidelines
- Links to software tools
- Single entry point on current data rescue projects
- Tips and tricks from past projects
- Enabling collaboration among organizations, development agencies, donors, scientists NGOs, etc.
- Accelerate data rescue in support of climate assessment and adaptation

The screenshot shows the homepage of the DARE portal. At the top, there is a navigation bar with logos for the European Union, Copernicus, and Climate Change Service, along with links for Home, About, Tools, Science, Inventories, Data sharing, and News. Below the navigation bar, there is a large banner image showing a server room with a cart in the foreground. The banner text reads "Data Rescue Tools" and "Data rescue guidelines and best practice instructions". Below the banner, there is a section titled "Range of software and technical tools for data rescuers from entry to expert levels". To the right of the banner, there is a "News" section with a headline "Upcoming presentation DARE Portal at EMS, Barcelona" and a date "2024-09-26". The news section includes a brief description of the portal and a link to the abstract.

implemented by ECMWF as part of The Copernicus Programme

Login Registration English

Home About Tools Science Inventories Data sharing News

Data Rescue Tools

Data rescue guidelines and best practice instructions

Range of software and technical tools for data rescuers from entry to expert levels

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

2024-09-26

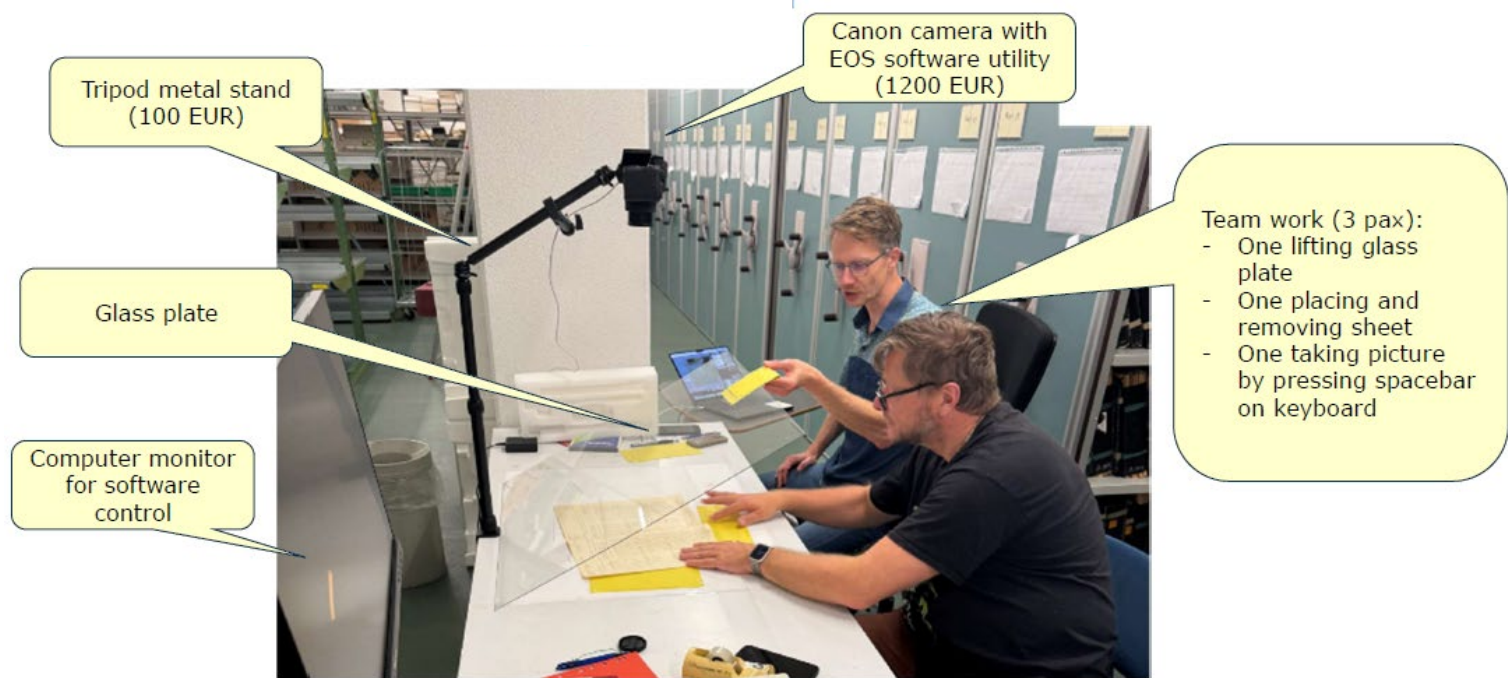
The new data rescue portal will be presented at the EMS Annual Meeting in the session on [Climate Monitoring data rescue, management, quality and harmonization](#) on Friday afternoon the 6th September. Read the [abstract](#) on the website of the EMS.

Marijas van der Schuer will present on the changes from the test portal and she will...

Step 2 – Digitize sheets

- 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

(Set-up used during innovation weeks at KNMI)



Digitization within IPDC at Aruba, Curaçao and St. Maarten

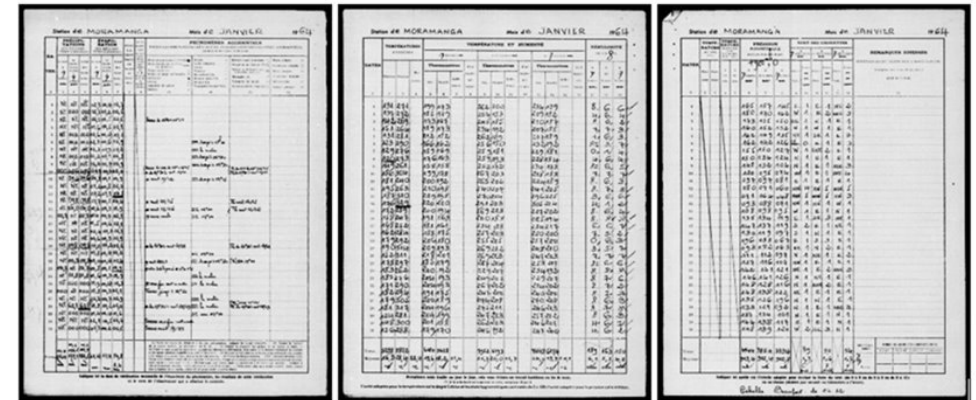
- International Panel on Deltas and Coastal Areas (IPDC) made budget available for digitization for the NMSs Aruba, Curaçao and St Maarten.
- Inventories are made of their physical archive
- KNMI provided guidance during this project to the NMSs on i.e. the image set-up in Aruba and Curaçao
- Annotations with and without AI are planned within this project.



Step 3 – Safeguard digital images

- Build from the ACMAD images (approx 200 TB) already secured under precursor C3S activities in collaboration with RMI in the first C3S period.
- Offer to international data rescue community to host additional images prioritising those with images at risk of loss
- The repository will be developed with archivists and use tagging for aspects: geolocation, time, transcription status, duplicates (and best duplicate indicator) that will aid navigation.

(KNMI is involved in this C3S2 task to set-up the image repository)



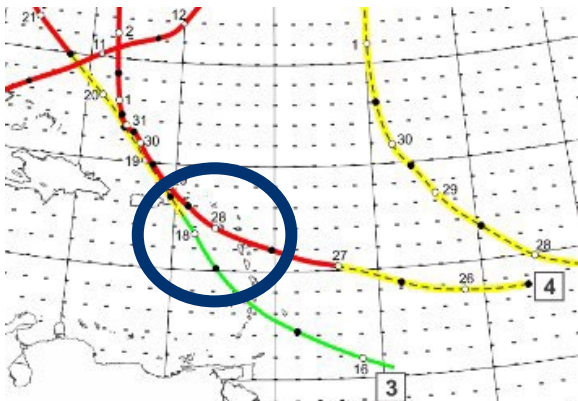


Step 4 – Annotation - option 1: manual

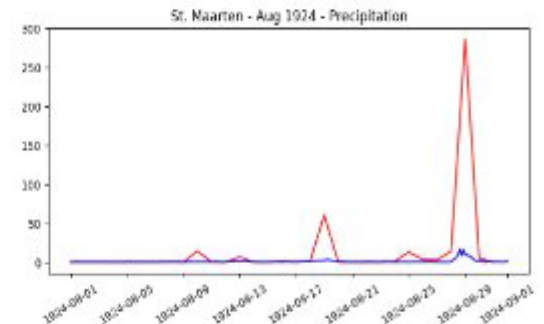
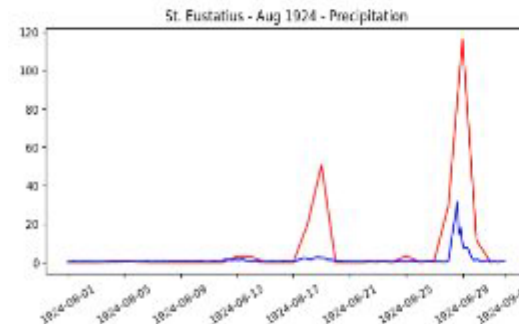
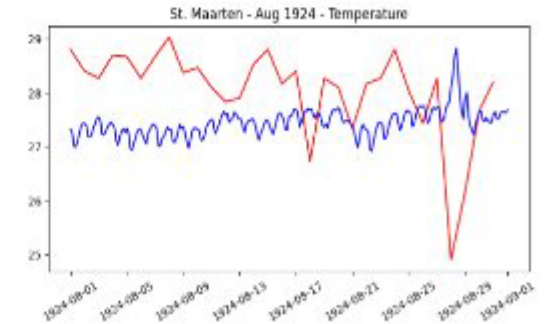
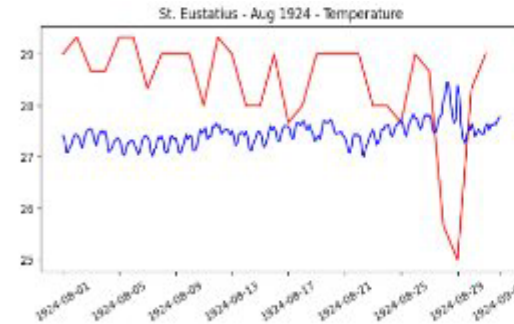
1. Contracting local people
2. Students:
 - Part of Curriculum:
 - Hands-on exercise
 - Learn useful transferrable skill
 - Feeling of a worthwhile contribution to global project
 - Example of classroom materials: <https://datarescue.climate.copernicus.eu/classroom-materials-climate-data-rescue-africa-project-clidar-project> (Noone et al., 2004)
 - Useful paid work and possibility for them to learn from work at Meteo Service:
3. Contracting people from in other countries via online platforms:
 - Experience for 4 sheets from St. Eustatius (via <https://www.upwork.com>):
 - Job was picked up in 5 hours, completed in 48 hours. \$10 dollars were paid (soft bargaining)

Example: 20CRv3

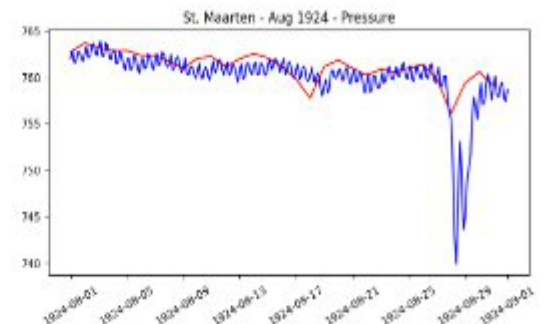
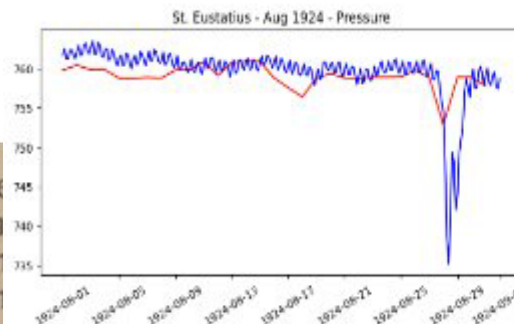
- Identified hurricane: August 1924
- 20CR grids are too large to represent the smaller BES-islands
- Difference with 20CRv3:
 - Temperature is overestimated
 - Precipitation underestimated
 - Pressure underestimated in observations



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



Step 4 – Annotation - option 2: Citizen Science

- Using Zooniverse
- Workflow established for 1911 precipitation
- The output can be used for climate science, AI/ML
- <https://www.zooniverse.org/>



Link to ACMAD collection on zooniverse

Weather Archive Africa ✓

Research

RESEARCH

THE TEAM

EDUCATION

FAQ

The ACMAD collection is a collection of c.4 million images of historical weather data from 43 African countries. Originally rescued by the Royal Meteorological Institute of Belgium (RMI) and the World Meteorological Organisation working in partnership with African Centre of Meteorological Applications for Development (ACMAD) in the late 1980s. These were imaged and saved onto microfilm and microfiches. Many of the original paper documents are believed to have been destroyed after this process and the microfilm and microfiche were feared to have degraded over time due to poor storage conditions. However, copies of the images were found on micro-film in a climate controlled cabinet in RMI. Further rescue efforts were carried out as priority in 2021 by the European Union's Copernicus Climate Change Service to save the images onto a tape archive, therefore preventing the data from being lost forever from the decaying film and fiche. The Copernicus Climate Change Service holds these data and has the agreement to rescue and make available these records to the research community. But first we need to catalogue and understand the quality of the images and this is where you come in.

Country	Date Coverage	Station Count	Country	Date Coverage	Station Count
Angola	1891-1949	14	Algeria	1891-1991	109
Benin	1899-1994	11	Madagascar	1899-1990	14
Botswana	1950-1992	14	Kenya	1894-1990	108
Burkina Faso	1900-1990	109	Tanzania	1900-1990	78
Burundi	1900-1990	92	Senegal	1894-1990	104



Zooniverse: Create your own workflow

PROJECT #26806

View project

Project details

About

Collaborators

Field guide

Tutorial

Media

Visibility

Talk

Data Exports

Workflows

Subject Sets

Translations

NEED SOME HELP?

Read a tutorial

Ask for help on talk

Ratings

DEVELOP 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 task will appear as an option on the project front page.

TASKS

How many mm 'Neerslag' are recorded on 'Datum' 1 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 2 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 3 in the FARM

Which 'maand' is recorded?

How many mm 'Neerslag' are recorded on 'Datum' 4 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 5 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 6 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 7 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 8 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 9 in the FARM

How many mm 'Neerslag' are recorded on 'Datum' 10 in the FARM

What has been the 'som' of 'Neerslag' from day 1-10?

Does the header of the last column state 'The Farm' or 'Pleasure'?

Which year is recorded - please, indicate the full year (18XX) instead

Add a task

MAIN TEXT

Which year is recorded - please, indicate the full year (18XX)

Describe the task, or ask the question, in a way that is clear to a non-expert. Note: this is a label for the task; 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 highlight portions of their transcription.

☐ Deletion

☐ Insertion

☒ Unclear

NEXT TASK

Which 'maand' is recorded?

Delete this task

FIRST TASK

Which year is recorded - please, indicate the full year (18XX)

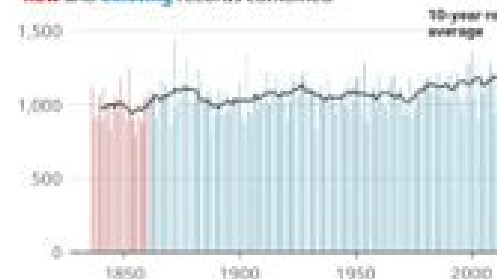
11

Citizen Science 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 1960s, 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.

Step 4 – Annotation – Option 3: AI/ML

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

<https://datarescue.climate.copernicus.eu/aiml-and-data-rescue>

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 $69\% \pm 15\%$ ($n = 920$). The 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 associated 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 [outpost](#) on this experiment.

[Zheng et al. \(2024\)](#) shows that their preliminary workflow offers a foundation for future attempts to automate the transcription of historical records, although there is still much to consider as to what the path forward will be. The python scripts are available on [GitHub](#).

date	val1	val2	val3	val4	val5	val6	val7	val8	val9	val10	val11	val12
1869	27	01	187	242	763	76						
2 days	18870.8	8893.5	18.3	2.8	7.8	7.6						
3 days	17405.33	4438.4	1480.1	2.5	0.8	2.8						
4 days	14938.8	10042	888.7	7	2	2.8						
5 days	1481.8	1481.8	1481.8	9.8	9.8	9.8						

Example AI/ML

Table recognition

LICHTDRUKING van Nieuw op O.P.C.			TEMPERATUUR °C.						DAMPDRUKING mm.				BETENBAARDE VOCHTHOUD gram.				WINDRICHTING EN KRACHT n. per m.m.f.			REUKEN D.-N.				
			7.30		13.30		19.30		Binn. 7.30-13.30 van 16.30 tot 19.30		Maa- tem.	Maa- norm.	7.30	13.30	19.30	Binn. 7.30-13.30 van 16.30 tot 19.30		7.30	13.30	19.30	7.30	13.30	19.30	
7.30	13.30	19.30	damp- druk	tem- per.	damp- druk	tem- per.	damp- druk	tem- per.	damp- druk	tem- per.														
22.02	22.02	22.02	25.6	23.2	28.2	24.6	24.6	23.2	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.05	22.05	22.05	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.08	22.08	22.08	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.11	22.11	22.11	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.14	22.14	22.14	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.17	22.17	22.17	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.20	22.20	22.20	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.23	22.23	22.23	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.26	22.26	22.26	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.29	22.29	22.29	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.32	22.32	22.32	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.35	22.35	22.35	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.38	22.38	22.38	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.41	22.41	22.41	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.44	22.44	22.44	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.47	22.47	22.47	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.50	22.50	22.50	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.53	22.53	22.53	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.56	22.56	22.56	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
22.59	22.59	22.59	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.02	23.02	23.02	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.05	23.05	23.05	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.08	23.08	23.08	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.11	23.11	23.11	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.14	23.14	23.14	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.17	23.17	23.17	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.20	23.20	23.20	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.23	23.23	23.23	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.26	23.26	23.26	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.29	23.29	23.29	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.32	23.32	23.32	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.35	23.35	23.35	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.38	23.38	23.38	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.41	23.41	23.41	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.44	23.44	23.44	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.47	23.47	23.47	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.50	23.50	23.50	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.53	23.53	23.53	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.56	23.56	23.56	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
23.59	23.59	23.59	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.02	24.02	24.02	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.05	24.05	24.05	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.08	24.08	24.08	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.11	24.11	24.11	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.14	24.14	24.14	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.17	24.17	24.17	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.20	24.20	24.20	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.23	24.23	24.23	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.26	24.26	24.26	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.29	24.29	24.29	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.32	24.32	24.32	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.35	24.35	24.35	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.38	24.38	24.38	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.41	24.41	24.41	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.44	24.44	24.44	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.47	24.47	24.47	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6	22.6	58	22	17	24.4	5	2	2	6	
24.50	24.50	24.50	25.2	23.6	28.6	24.6	24.6	23.6	78.2	22.2	22.8	21.8	24.2	24.6</										

Optical Character Recognition

nr	LUCHTDRUKUNG mm. barid op 0° C.			TEMPERATUUR °C.						DAMPDRUKING mm.				REHURRELRE. VOCHTIGHEID pct.				WINDRICHTING EN KRACHT in per uur p'			BEWOL 0-10					
	7.30			13.30		19.30		Sun 7.30, 13.30 on 19.30		Maxi. max.	Mini. min.	7.30 11.30 19.30			Sun 7.30, 13.30 on 19.30			7.30	13.30	19.30	7.30	13.30	19.30			
	7.30	13.30	19.30	droog hul.	natte hul.	droog hul.	natte hul.	droog hul.	natte hul.			7.30	11.30	19.30	7.30	13.30	19.30									
1	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
2	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
3	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
4	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
5	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
6	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
7	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
8	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
9	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
10	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
11	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
12	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
13	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
14	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
15	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
16	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
17	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
18	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
19	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
20	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
21	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
22	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
23	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
24	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
25	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
26	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
27	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
28	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
29	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
30	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
31	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
32	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
33	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
34	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
35	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
36	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
37	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
38	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
39	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
40	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
41	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
42	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
43	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
44	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
45	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
46	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
47	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
48	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
49	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
50	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
51	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
52	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52	19	2.44	5.2	1.10	2	12	12	9	6
53	78.2	78.2	78.2	25.6	27.7	28.8	28.6	28.2	22.2	78.2	22.2	21.8	22.6	22.6	62.6	55	52									



Example AI/ML 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

Conclusion:

- Methods have not been proven to be good enough to be used with the high accuracy that our community is looking for.
- Advances in this field are rapid.

(Master student at KNMI is working on this)

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		

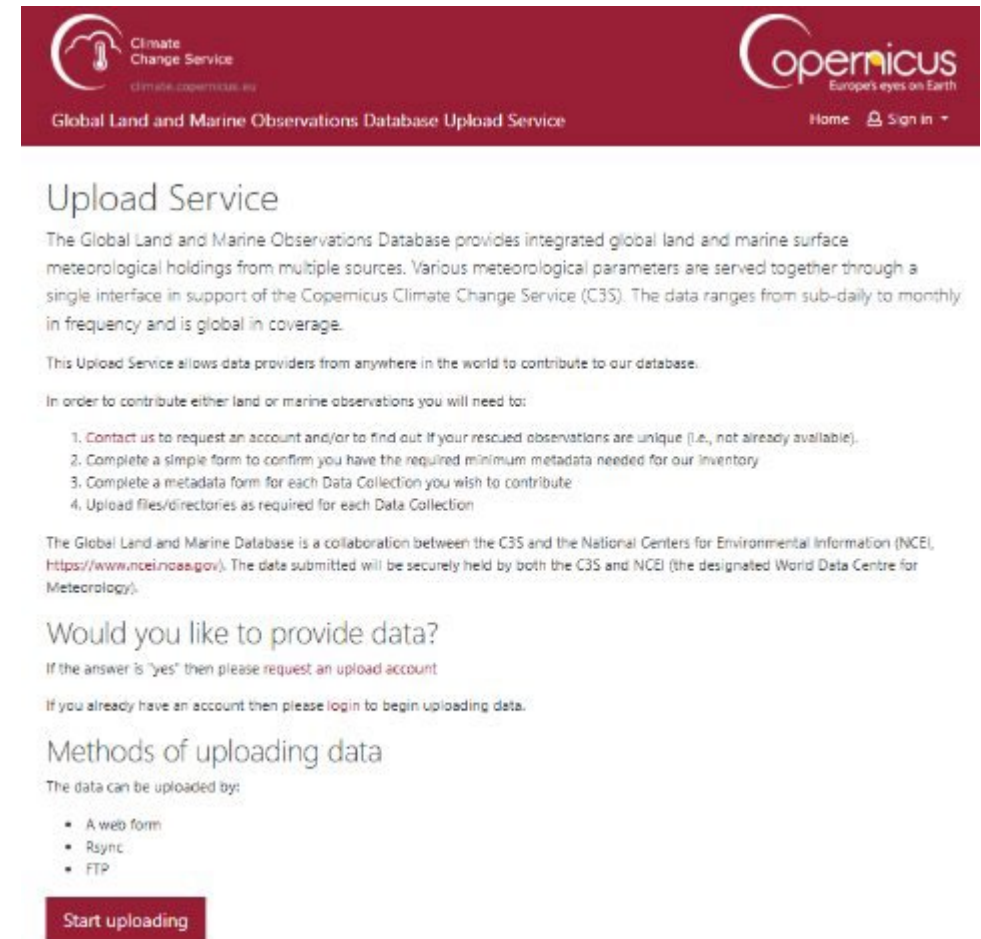
Step 5 – make rescued data available

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/>
(Maintained in C3S with KNMI as WP lead)

Data for Europe can also be shared with ECA&D
<https://www.ecad.eu>
eca@knmi.nl



The screenshot shows the 'Global Land and Marine Observations Database Upload Service' page. At the top, there are logos for the Climate Change Service and Copernicus. The main heading is 'Upload Service'. Below this, a paragraph explains that the database provides integrated global land and marine surface meteorological holdings from multiple sources. A sub-section titled 'This Upload Service allows data providers from anywhere in the world to contribute to our database.' is followed by a list of four steps: 1. Contact us to request an account, 2. Complete a simple form to confirm you have the required minimum metadata, 3. Complete a metadata form for each Data Collection, and 4. Upload files/directories as required. A note mentions the collaboration with NCEI. Further down, there are sections for 'Would you like to provide data?' with instructions on how to request an account or login, and 'Methods of uploading data' which lists web form, Rsync, and FTP. A red 'Start uploading' button is at the bottom.



PROGRAMME OF
THE EUROPEAN UNION



IMPLEMENTED BY



#EUSpace

Thank you



© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Slide xx: [element concerned](#), source: [e.g. Fotolia.com](#); Slide xx: [element concerned](#), source: [e.g. iStock.com](#)