PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

Climate Change Service climate.copernicus.eu

IMPLEMENTED BY ECMWF

#EUSpace

Facilitating data rescue initiatives

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

KNMI, The Netherlands

Introduction – Data Rescue

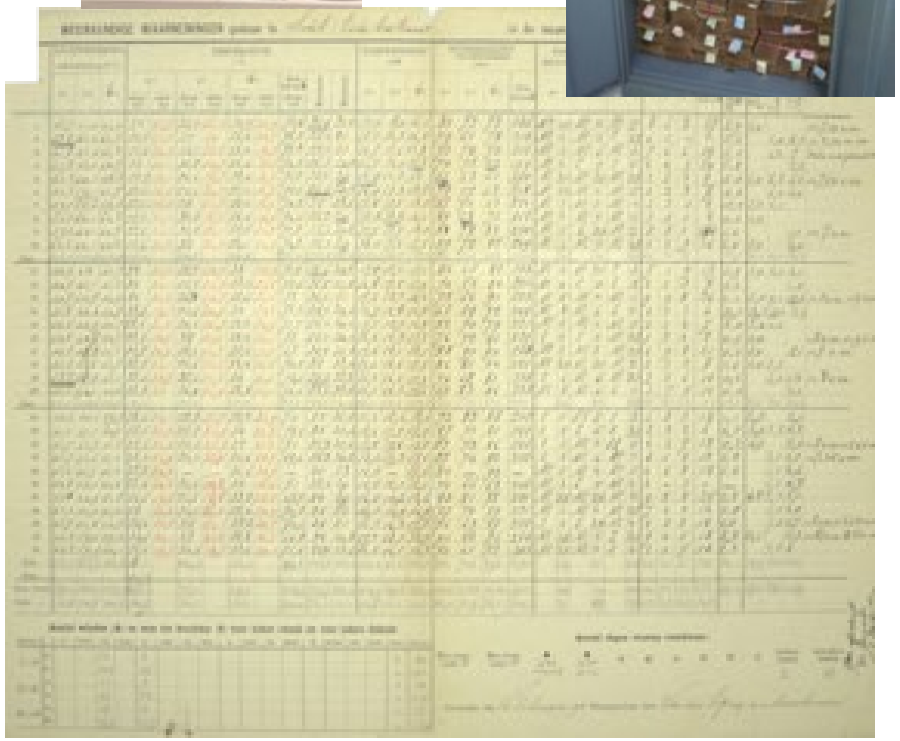
Extension of problem:

- 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

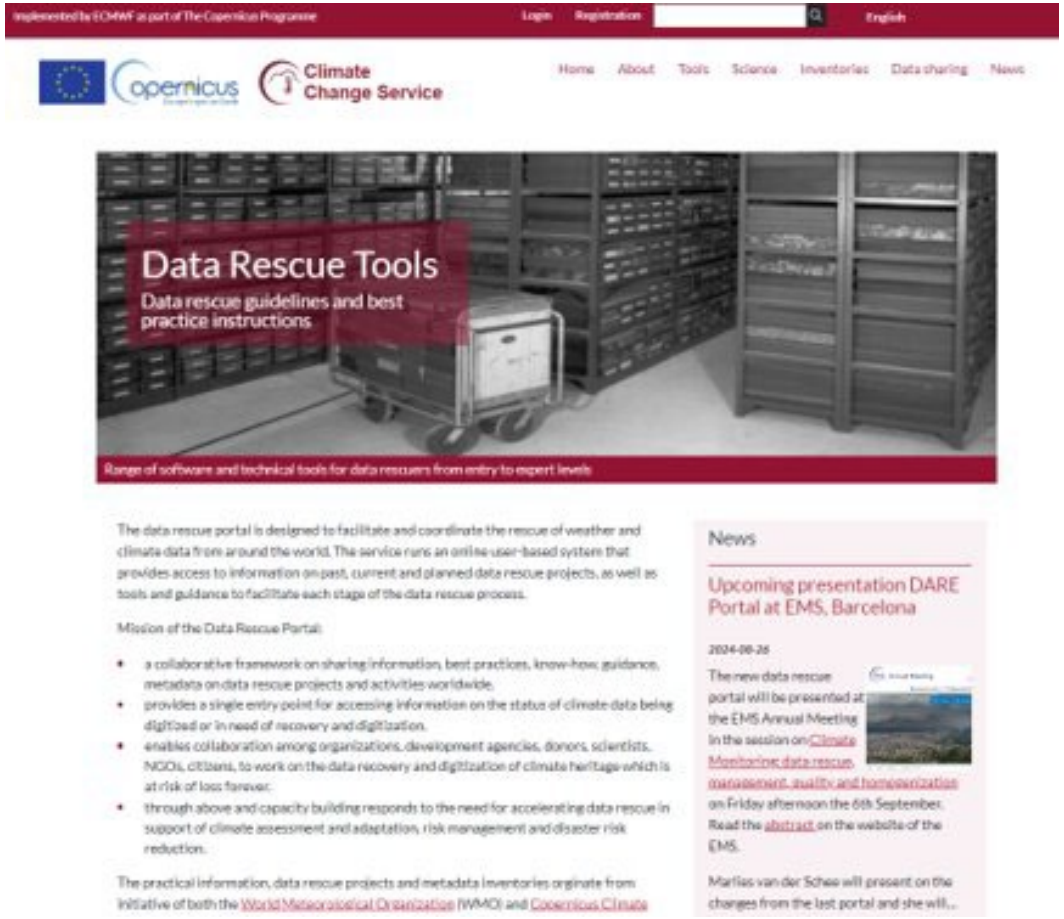
Share your project, ensures rescuing data only once!

- 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)

Step 1 – Data Rescue 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

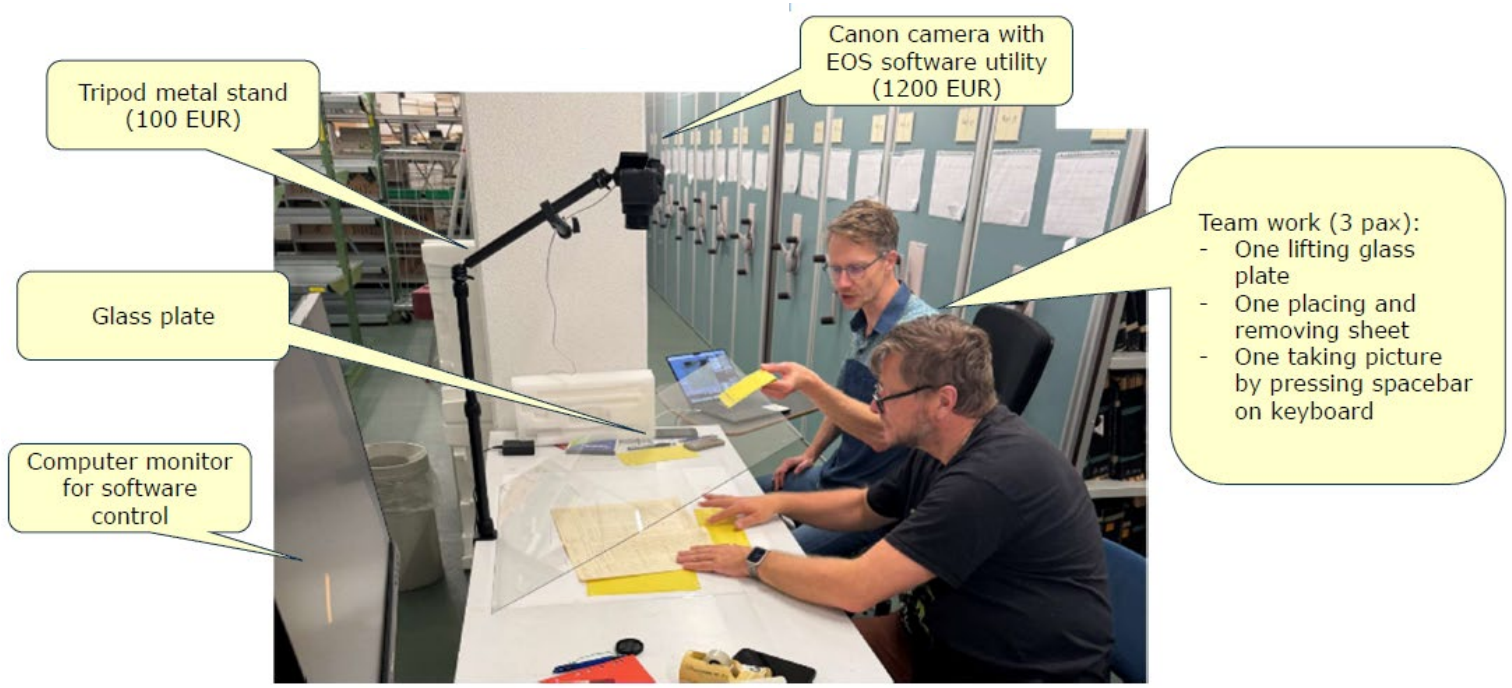
Homepage DARE portal



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)



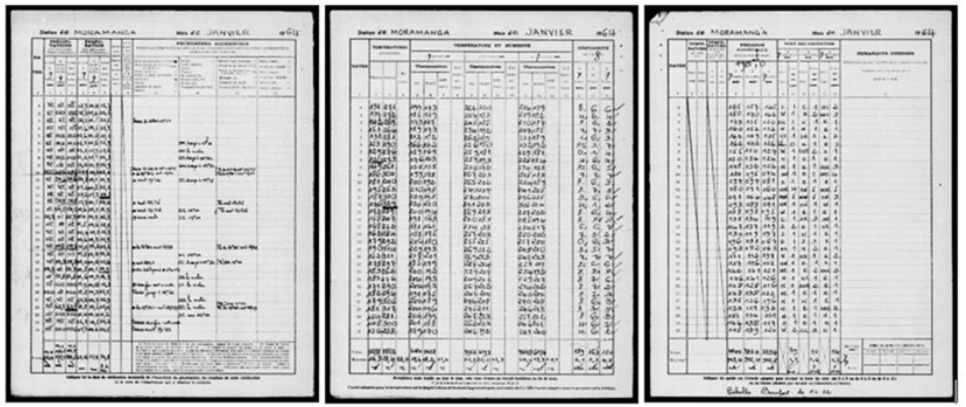
Digitization within IPDC at Aruba, Curaçao and St. Maarten

- 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)

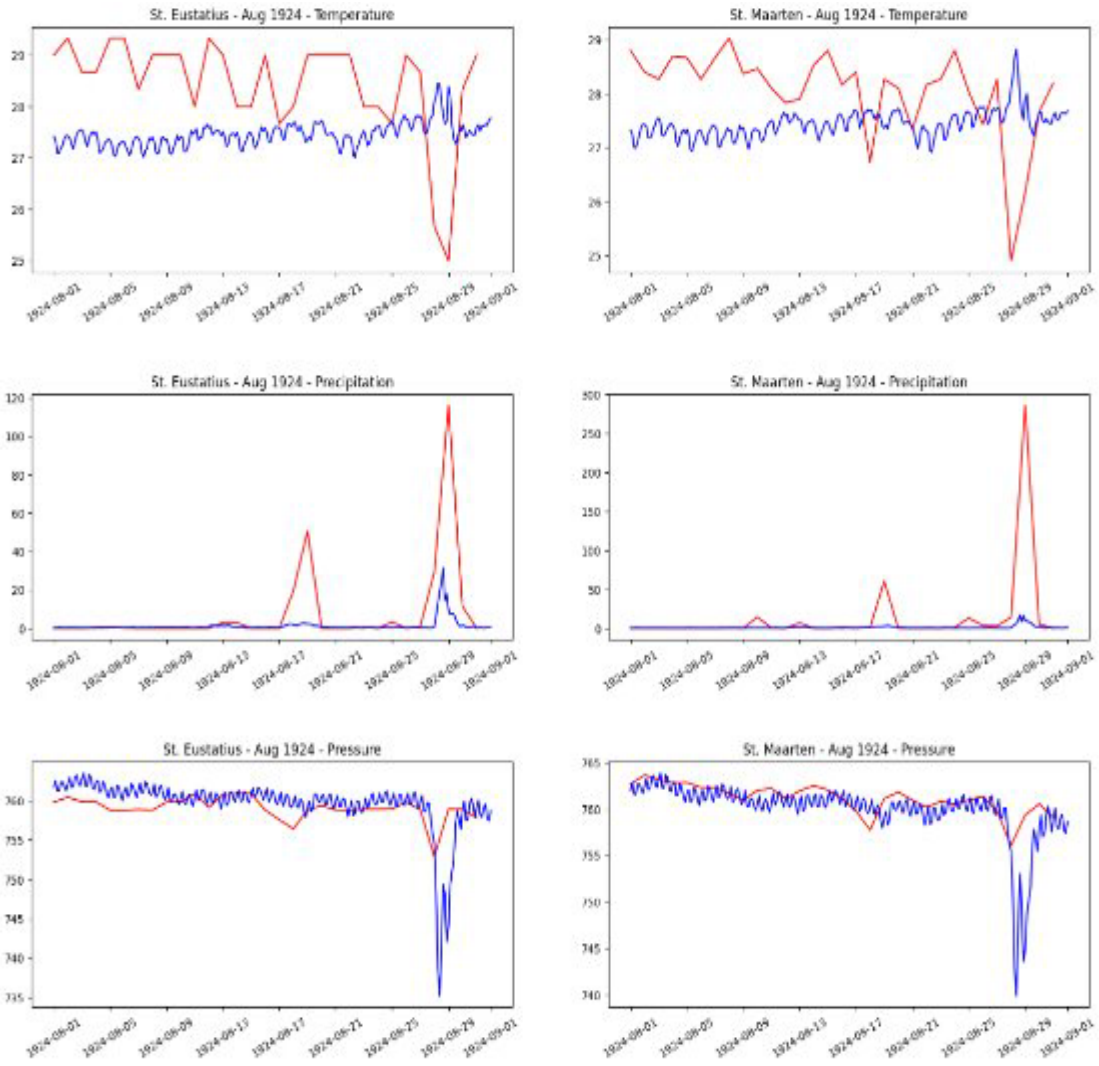


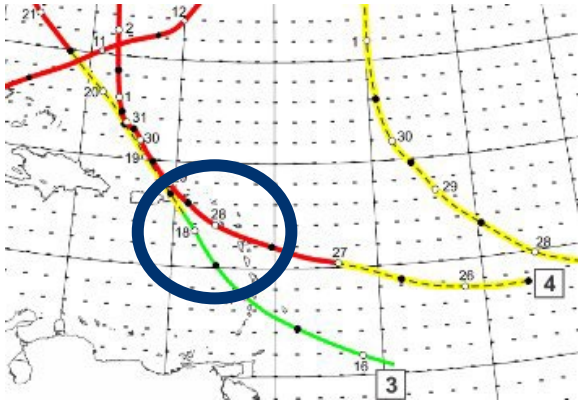
Step 4 – Annotation - option 1: manual

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

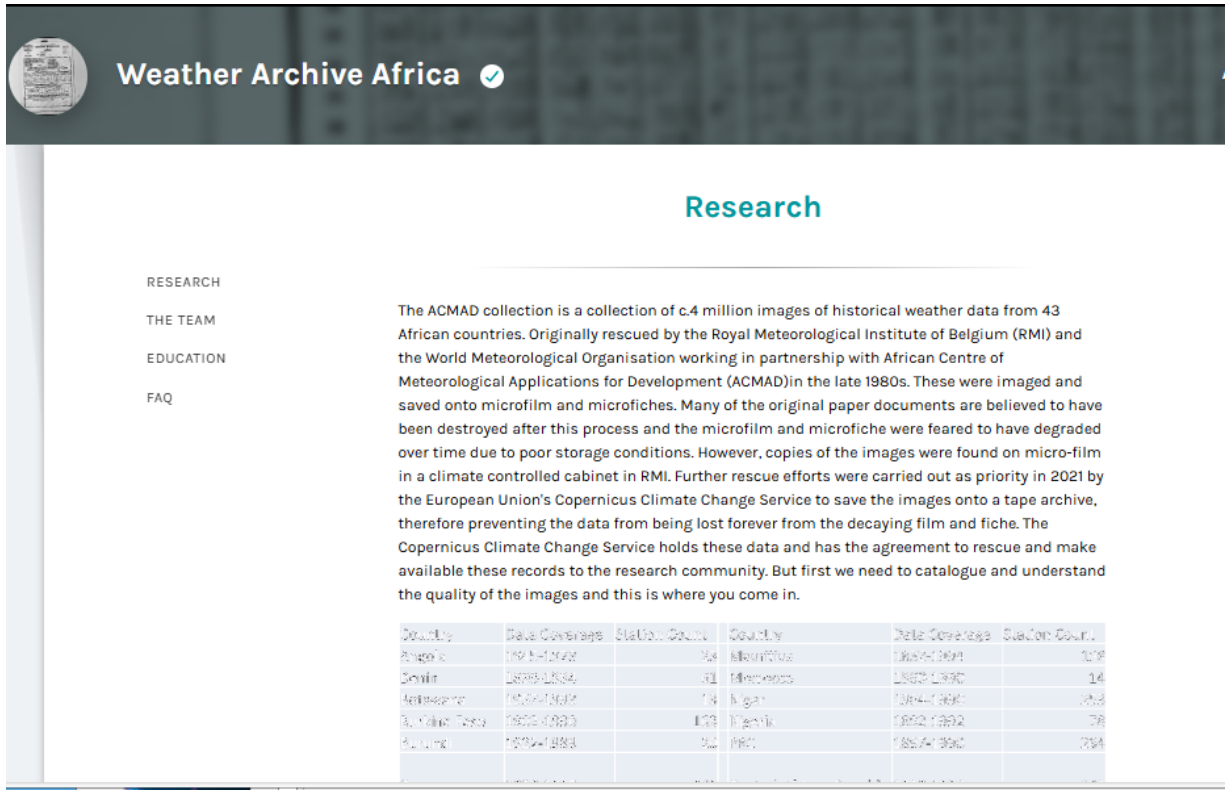





Orkaan.
De orkaan die verleden week hie
geseind was en in westelijke richtin
tusschen de eilanden Dominica en An
tigua gepasseerd is, heeft ook aan on
ze Bovenwindsche eilanden een be
zoek gebracht en volgens bericht van
den Gezaghebber ontvangen er nog
al aardig huisgehouden.

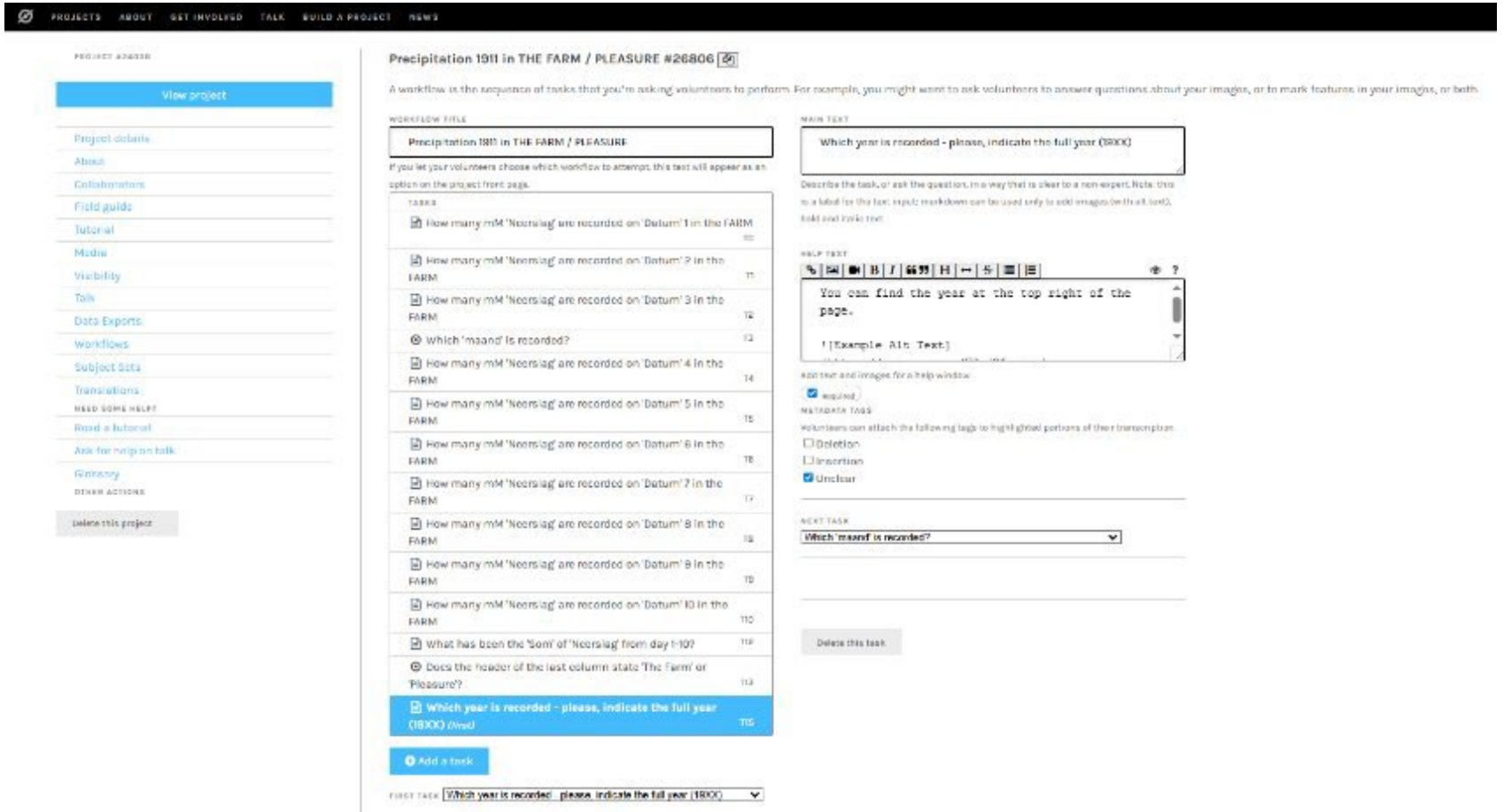
Step 4 – Annotation - option 2: Citizen Science

- 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

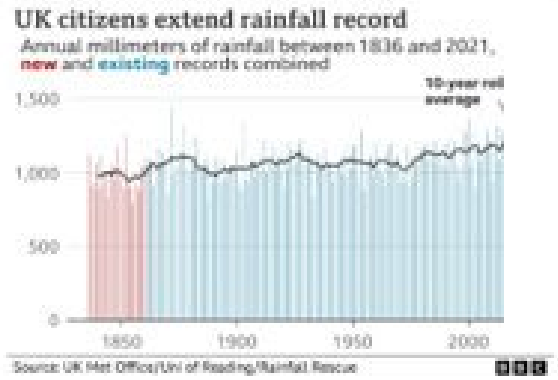
Zooniverse: Create your own workflow

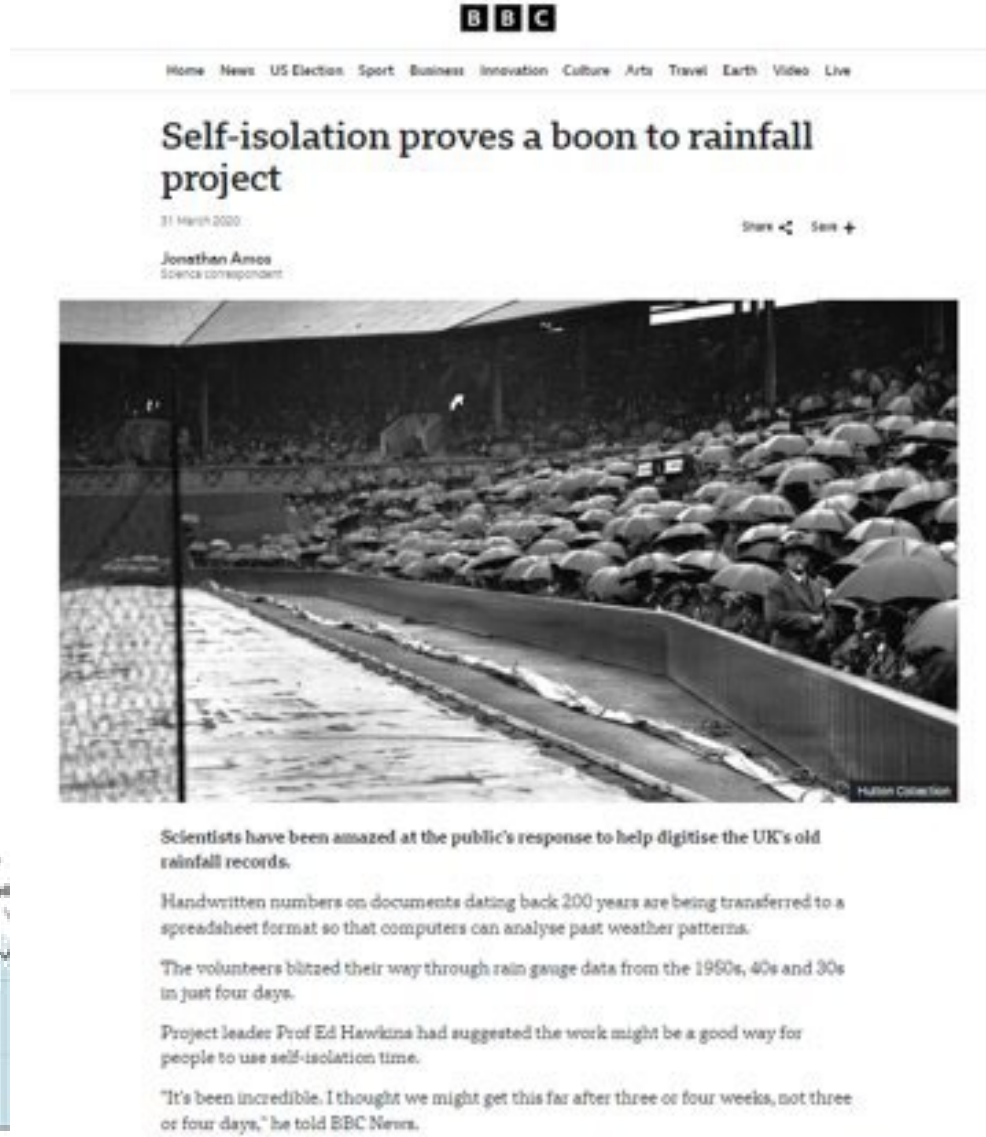


Citizen Science example

- UK MetOffice/University of Reading
- Covid lockdown

- 16,000 volunteers
- 65,000 scanned pieces
- 5.4 million rainfall observations




BBC

Home News US Election Sport Business Innovation Culture Arts Travel Earth Video Live

Self-isolation proves a boon to rainfall project

31 March 2020

Share Save

Jonathan Amos
Science correspondent

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.

Source: https://www.bbc.com/news/science-environment-52040825

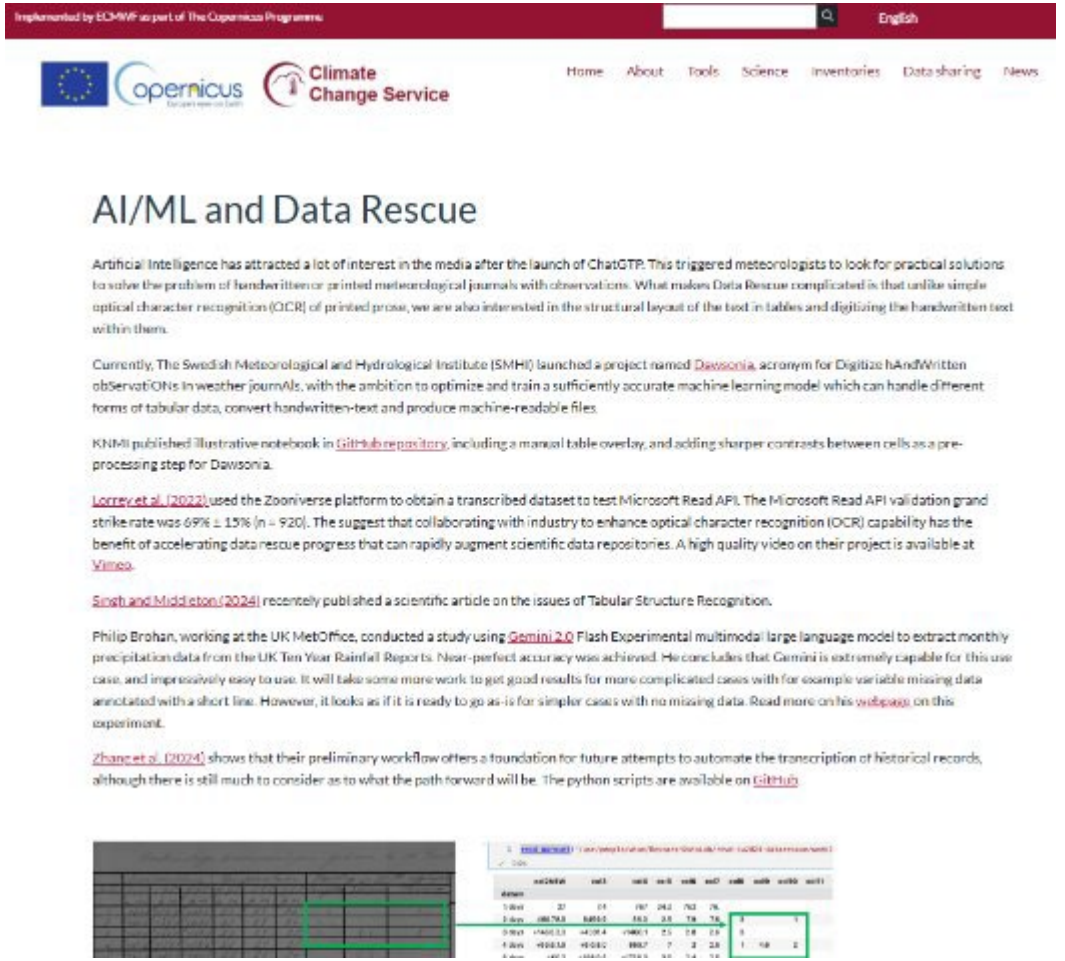
Step 4 – Annotation – Option 3: AI/ML

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

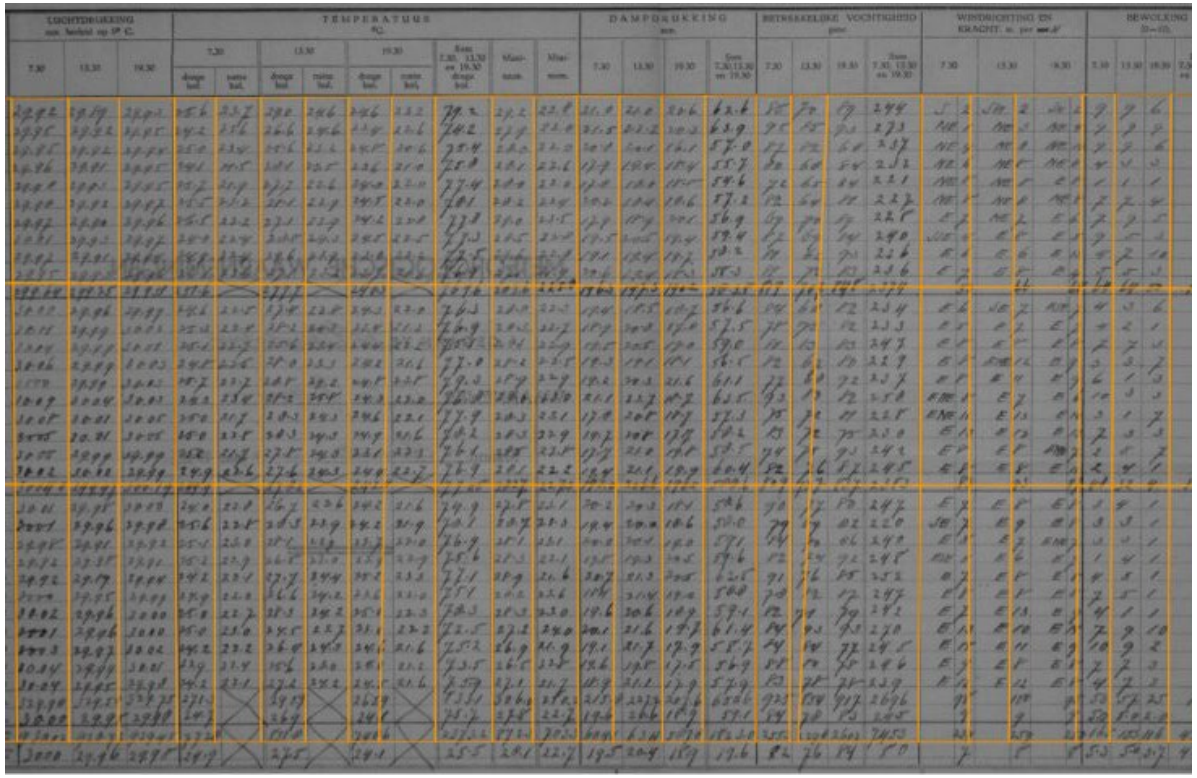
Literature review on Copernicus website

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

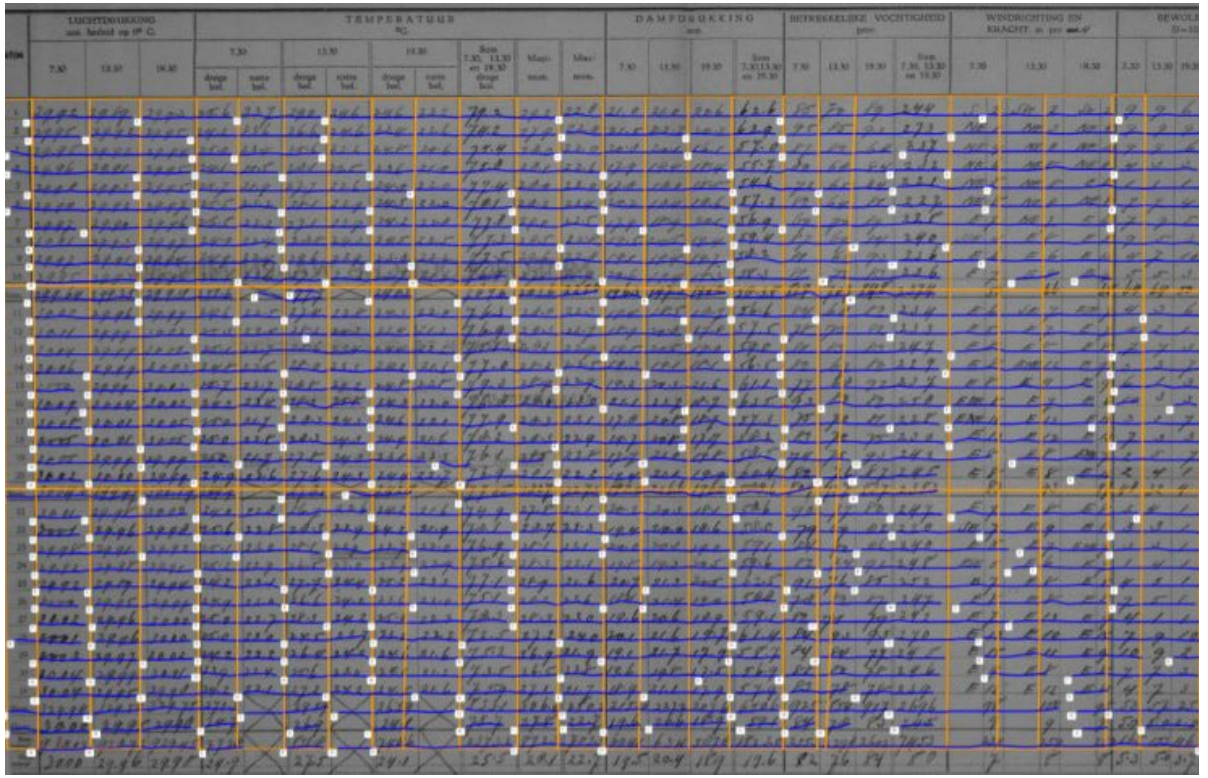


Example AI/ML

Table recognition



Optical Character Recognition



Example AI/ML MeteoSaver v1.0

- 74% matches the manually transcribed record

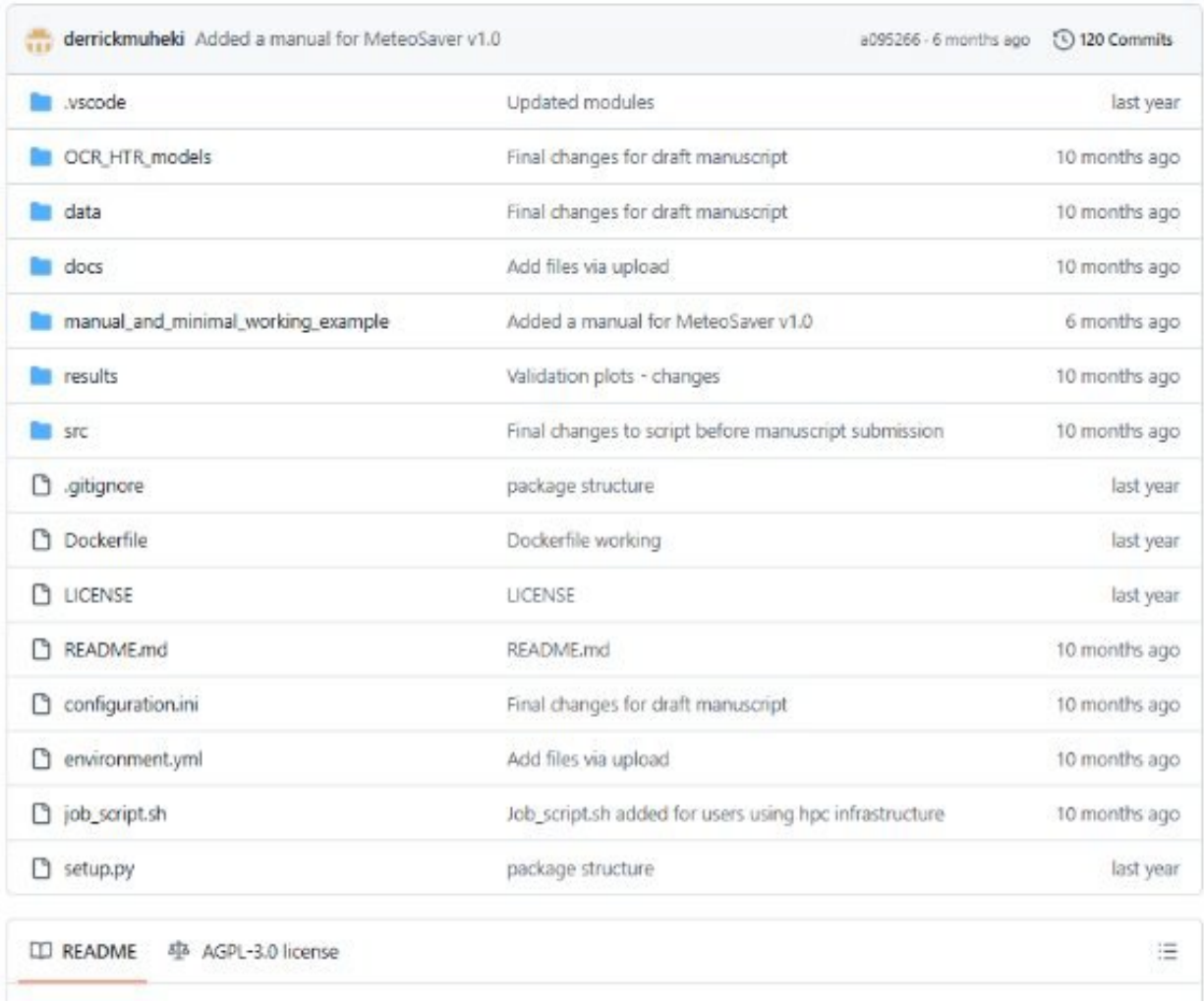
Recommendation:

- Enhace 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)



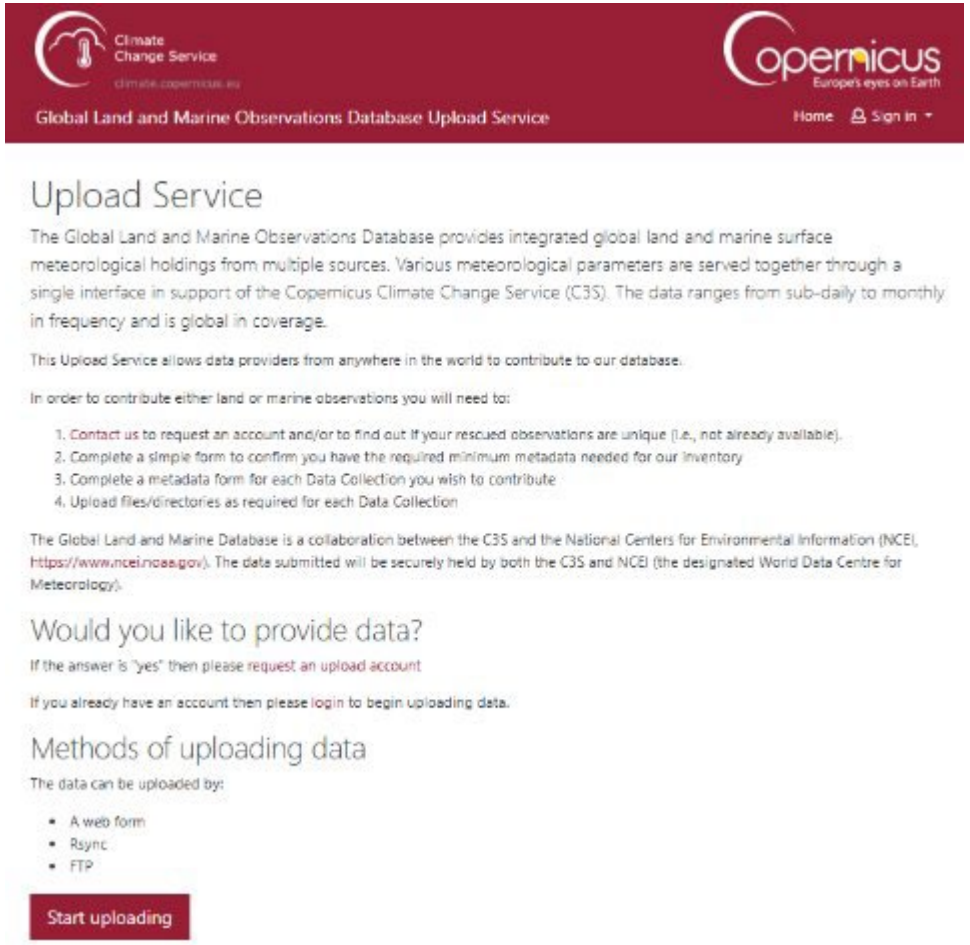
Step 5 – make rescued data available

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



PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

Climate Change Service climate.copernicus.eu

IMPLEMENTED BY ECMWF

#EUSpace

Thank you