

RESEARCH ARTICLE

Digitizing historical daily weather bulletins through citizen scientists: The ReData project

Alessandro Ceppi^{1,2}, Veronica Manara^{3*}, Yuri Brugnara^{4,5}, Gabriele Buccheri³, Goffredo Caruso¹, Luca Cerri², Maria Di Giovanni³, Marco Giazzi², Ludovico Lapo Luperi^{6,7}, Luca Ronca², Elisa Sogno³, Maurizio Maugeri³

1 Department of Civil and Environmental Engineering (DICA), Politecnico di Milano, Milano, Italy, **2** Associazione Meteonetnetwork OdV, Milano, Italy, **3** Department of Environmental Science and Policy, Università degli Studi di Milano, Milano, Italy, **4** Institute of Geography, University of Bern, Bern, Switzerland, **5** Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland, **6** CEA - Paris-Saclay, Saclay, France, **7** INFN - Sezione di Milano, Milano, Italy

* veronica.manara@unimi.it



Abstract

In recent decades, numerous climate data rescue programs have begun in many countries worldwide. These projects aim to preserve data recorded on paper sheets, which are vulnerable to deterioration, and to make them accessible to the scientific community. This enhances the accuracy of climatological studies and historical reconstructions, including those focused on specific events. This study presents the designed framework developed in the ReData (Recovery of Data) project, launched by the Meteonetnetwork association in collaboration with the University of Milan in 2017, and upgraded in 2024 on the Zooniverse platform, which engages volunteers in scientific research activities; in particular, it showcases the methodology implemented for this climate data rescue initiative. The job leverages the potential of citizen science to digitize meteorological data, collected by the Italian Royal Central Meteorological Office (RCMO) from 1879 to 1940 and published in daily meteorological bulletins, on a platform specifically designed to facilitate large-scale digitization and ensure user accessibility. In addition, as a practical application of digitized data, a case study is presented involving a synoptic reconstruction of the flood event that affected the River Adige in northeastern Italy in September 1882. The overall project provides critical data for reanalysis models, and it enhances the understanding of historical climate trends over the Italian peninsula, offering significant cultural and scientific value.

OPEN ACCESS

Citation: Ceppi A, Manara V, Brugnara Y, Buccheri G, Caruso G, Cerri L, et al. (2026) Digitizing historical daily weather bulletins through citizen scientists: The ReData project. *PLOS Clim* 5(3): e0000865. <https://doi.org/10.1371/journal.pclm.0000865>

Editor: Alfonso Fernández, Universidad de Concepción: Universidad de Concepcion, CHILE

Received: September 18, 2025

Accepted: February 19, 2026

Published: March 24, 2026

Copyright: © 2026 Ceppi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The scanning of the daily RCMO weather bulletins are freely available as JPG files (https://doi.org/10.13130/RD_UNIMI/R1GVKF).

Funding: This work has been financed by Research Funds from the Italian Ministry for University and Research (PRIN 2022

1 Introduction

Climate change is one of the most urgent and complex challenges in the last decades [1]. Historical data allow scientists to detect long-term trends, identify anomalies, improve reanalysis and validation of climate models, which are fundamental for

– CN4RWK–CCHP–ALPS Climate Change and Hydro–Power in the Alps, funded by the European Union (Programme Next Generation EU)).

Competing interests: The authors have declared that no competing interests exist.

developing effective mitigation and adaptation strategies. Trends in meteorological variables can be studied by investigating records of observational data at selected station sites (see, e.g., [2] for Italy or [3] for the Greater Alpine Region) as well as by considering gridded datasets of observational records (see, e.g., [4] for Europe, [5] worldwide). In addition, global reanalysis datasets, such as the ECMWF (European Centre for Medium-Range Weather Forecasts) Reanalysis v5 (ERA5, [6]) and the NOAA Twentieth Century Reanalysis (20CR) [7], play a key role in reconstructing climate of the past by assimilating old observations into state-of-the-art dynamic models [8]. However, both for gridded datasets of observational data and for reanalyses, the reliability of reconstructed fields directly depends on availability, homogeneity, and spatial distribution of the underlying measurements.

Unfortunately, many old meteorological data, originally recorded on paper documents, have yet to be recovered, and their preservation is essential for several reasons [9]. First, these paper records are often significantly deteriorated, posing a serious risk of permanently losing valuable information; moreover, making these data accessible for climatological research would enhance the accuracy as well as the spatial and temporal coverage of datasets used to investigate the climate of the past. The recovery of new data can therefore allow researchers, on one hand, to improve the climate reconstruction at local scale and, on the other hand, it can contribute to increasing the accuracy of global reanalysis datasets [10].

The preservation of historical meteorological data has been ongoing for several decades, such as the International Surface Temperature Initiative [11], but its progress varies depending on the region, accessibility of sources, and availability of resources. Several projects have already been completed, and others still need to be initiated to safeguard data that have not been digitized yet. At international level, the World Meteorological Organization (WMO) coordinates global efforts to standardize and digitize climate data, creating databases that are accessible to scientists worldwide. For instance, the Data Rescue Portal, by the Copernicus Climate Change Service (C3S), collects and disseminates information on past, current, and planned data rescue projects (<https://datarescue.climate.copernicus.eu/>, accessed on 1st July, 2025).

One of the largest climate-related citizen science data rescue projects ever completed comes from the Rainfall Rescue project [12] where more than 66,000 paper sheets containing 5.28 million handwritten monthly rainfall observations recorded across the UK and Ireland between 1677 and 1960 were digitized. This achievement was made possible by the contributions of over 16,000 volunteers who participated during the early months of the COVID-19 lockdown in 2020 through the Zooniverse platform. In New Zealand, the Southern Weather Discovery is another remarkable initiative focused on transcribing meteorological records from thousands of ships that crossed the Southern Ocean during the 19th and 20th centuries to improve the understanding of historical weather patterns and assess long-term climate trends [13]. In France, the national meteorological service Météo-France, in collaboration with the National Archives and supported by the BNP Paribas Foundation, has been engaged since 2012 in a program to preserve climate archives from both France and its former colonies; covering the period from 1850 to 1960 (<http://archivesduclimat.meteofrance.fr/>, accessed on 1st July, 2025).

Another indicator of the growing importance of climate data preservation within today's scientific community is the emergence of numerous non-profit organizations dedicated to promoting and raising awareness about this issue. A remarkable example is the International Environmental Data Rescue Organization (IEDRO), a U.S.-based non-profit society that collaborates with the WMO, the National Oceanic and Atmospheric Administration (NOAA), and the National Meteorological and Hydrological Services (NMHS) of developing countries to support data recovery efforts. Lastly, other worthy examples come from the potential of enrolling undergraduate students in climate data rescue projects [14–16], even for solving metadata matters [17].

In Italy, a country with a rich heritage of historical meteorological data, home to the invention of famous instruments such as Galileo's thermoscope and Torricelli's barometer, and creator of the first international observational network by the Accademia del Cimento in the 17th century [18], the Italian Association of Atmospheric Sciences and Meteorology (AISAM) plays a central role in national data rescue efforts. In 2022, AISAM launched the Cli-DaRe (Citizen Science for Italian Climate Data Rescue) project, aiming to initiate a long-term program that leverages the vast potential of citizen science to recover the extensive meteorological data still preserved in Italian paper archives. The first initiative under this project, Cli-Dare@School, involved high school students in the recovery of historical rainfall series from monographs published by the Italian Hydrographic Service between 1918 and 1966 [19–21]. Other projects are ongoing, and further information can be found at the following website: <https://aisam.eu/progetti-aisam/> (accessed on 1st July, 2025).

Under this framework, the proposed study focuses on a recent project regarding the recovery of historical climate data in Italy: ReData (<https://datarescue.climate.copernicus.eu/redata> accessed on 1st July, 2025). The project is led by the Meteonetwork (MNW) association [22] in collaboration with the University of Milan (UniMi). Initially launched in 2017 on MNW webpage (<https://www.meteonetwork.it/>, accessed on 1st July, 2025), it was later upgraded in 2024 and moved to the Zooniverse platform (www.zooniverse.org/projects/meteonetwork/redata, accessed on 1st July, 2025) to benefit from citizen scientists' contribution.

Citizen science involves the active participation of the public in scientific research, allowing individuals without formal scientific training to contribute to data collection, analysis, and problem-solving. This approach enables researchers to gather large volumes of data while fostering greater public engagement with scientific issues [23–25]. Organizations such as the Association for Advancing Participatory Sciences (AAPS, <https://participatorysciences.org/>, accessed on 1st July, 2025) and the European Citizen Science Association (ECSA, <https://www.ecsa.ngo/>, accessed on 1st July, 2025) have been established to support and promote these initiatives, with this latter providing a framework of key principles for best practices. Indeed, the digital age has significantly advanced citizen science, with dashboards like Zooniverse (<https://www.zooniverse.org/>, accessed on 1st July, 2025) enabling global collaboration.

In the ReData project, we scanned all the data contained in the daily weather bulletins issued by the Italian Royal Central Meteorological Office (RCMO) between December 1879 and April 1940. This Meteorological Office was established on the 26th of November 1876 (Royal Decree n. 3534), some years after the unification of the Kingdom of Italy (1861), and it played a pivotal role in coordinating meteorological observations across the country for a long time. It evolved from earlier initiatives at the Collegio Romano in Rome, where meteorological data were originally collected to support astronomical research, and set the basis to unify Italy's fragmented meteorological networks by standardizing measurements and enabling consistent data collection nationwide. Over time, the office's responsibilities expanded to include geodynamics, later evolving into the Central Office of Meteorology and Geophysics. Despite challenges, such as limited resources and shifting institutional priorities, the RCMO laid the groundwork for Italy's modern meteorological and seismic services. In 1925 the Italian government decided to move the forecasting activities to the Italian Air Force. The RCMO, therefore, shifted its focus toward climatology and agricultural meteorology, and the archives of this institution are currently part of the Italian Agriculture and Environment Research Centre (CREA).

The bulletins, issued by the Meteorological Office and scanned within the ReData project, report the weather analysis daily performed. They include everyday data for many Italian and some European stations as well as weather charts

displaying the spatial distribution of the most important atmospheric variables. Therefore, one of the main tasks of the ReData project, which is currently ongoing, is the digitization of the huge amount of data contained in these bulletins.

Based on a detailed analysis of the evolution of the RCMO weather bulletins, regarding both monitoring stations and recorded meteorological variables, the aim of this study is to introduce the implemented design and describe the methods for this activity, which is structured to streamline the digitization process for citizen scientists, even replicable by other researchers. The final goal is making freely accessible all the information contained in the RCMO weather bulletins to support and contribute to scientific research. In order to better highlight the potential of rescued data, a case study is presented, regarding an analysis of a severe flood event occurring in September 1882 in northeastern Italy, which, due to persistent heavy rainfall, caused huge damages, including the collapse of Verona's Ponte Nuovo bridge. The weather data from RCMO bulletins showed detailed information at synoptic scale over the Italian peninsula, demonstrating their added value in the meteorological event reconstruction.

The main goal of this paper is to highlight the methodology we adopted to engage citizen scientists for the data digitization activities carried on in the ReData project. The text is structured as follows: Section 2 describes the database, i.e., the daily weather bulletins edited by the RCMO. The core of the paper is presented in Section 3, which details how we organized the ReData digitization on the Zooniverse platform and outlines key challenges encountered in this activity. Lastly, a reconstruction based on sea level pressure data related to a severe flood that occurred in the Adige River basin is illustrated in Section 4, while discussion and conclusions are reported in Section 5.

2 The daily weather bulletin of the RCMO

In 1879, the RCMO established a national network for the collection of near real-time observations. Using telegraph technology, the station observations were transmitted to the Central Office in Rome, enabling the publication of a daily weather bulletin. At the beginning, the network included 10 Italian stations, which quickly increased, reaching about 70 within a few decades. Starting from 1914, it further expanded to also include, beside the traditional network of meteorological observatories mainly located in Italian large cities, stations of harbor authorities and cities in newly acquired territories, such as Trento and Trieste, as well as colonial locations like Tripoli and Benghazi in Libya, amounting to more than 100 stations ([Fig 1](#)). Over the years, the number of stations showed significant variability, with a decrease after 1922; nevertheless, the station density remained high enough to meet the requirements for synoptic meteorology.

In addition to the number of reported Italian stations, many other changes occurred in the bulletin over the 60-year period during which it was issued. These changes concern the format and organization of the bulletin, the number of pages, the reported variables and weather charts, the information (station data and synthetic reports) from other European countries (not reported in [Fig 1](#)), and the format of short-term weather forecasts, referred to as *presagi*, which was reported on the last page. More detailed information on the temporal evolution of the daily weather bulletin issued by the Italian RCMO is provided in [Fig 2](#).

The daily weather bulletins of the Italian RCMO are preserved in various archives, either as separated files or compiled into annual volumes. Although relevant changes occurred over time, the general structure remained largely consistent, and it can be divided into four main sections: (i) a table of meteorological observations from Italian stations, (ii) a table of meteorological observations from stations in other European countries, (iii) one or more weather charts, (iv) other information, including a short-term weather forecast.

The most relevant section for the current objectives of the ReData project is the table reporting measurements from Italian stations. In some periods, this table is reported onto a single page; in others, it spans over two or three pages. While this table underwent changes over the years, it kept a stable structure during the period from 1882 to 1913 where the following meteorological variables were recorded every morning ([Fig 3](#)): barometric pressure at sea level, 24-hour barometric tendency, air temperature, 24-hour air temperature tendency, wind speed and direction, sky condition, sea

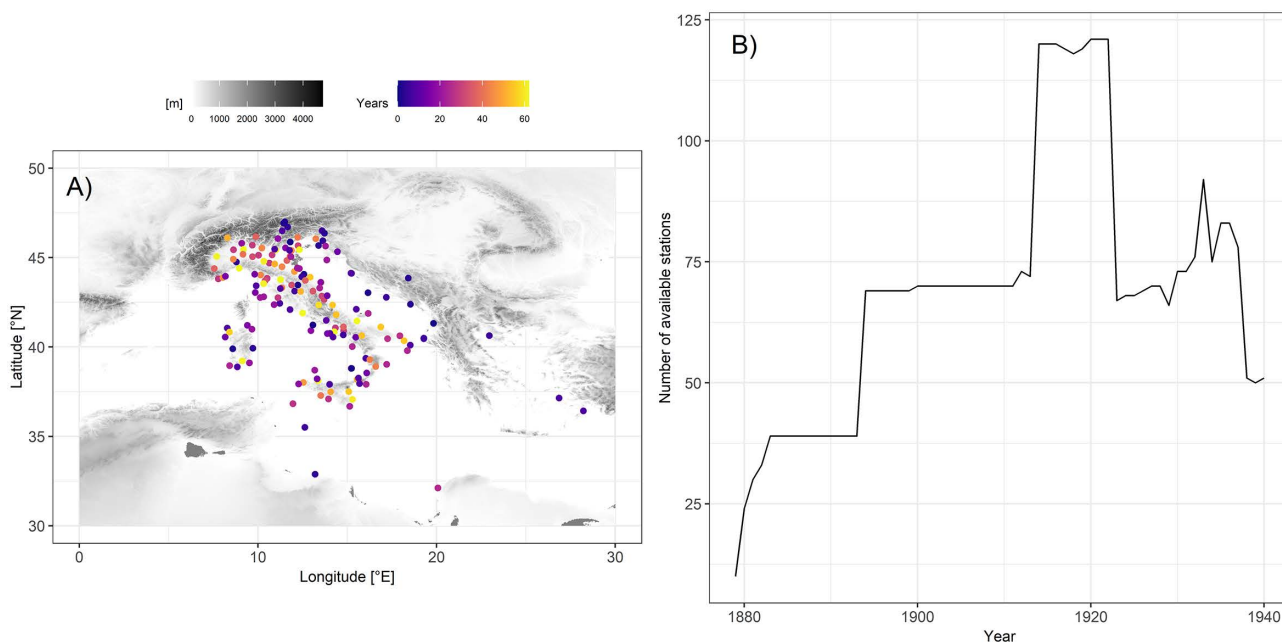


Fig 1. A) Spatial distribution of meteorological stations included in the RCMO weather bulletins where the color of the points represents the number of available years for each station. The orography of the region (USGS GTOPO30 digital elevation model - <http://www.usgs.gov>) is shown in grey colors. **B)** Temporal evolution of the number of available stations for each year.

<https://doi.org/10.1371/journal.pclm.0000865.g001>

state, cloud coverage, 24-hour rainfall, daily minimum and maximum temperatures, miscellaneous observations (typically, a brief comment on weather events over the previous 24 hours), and notes.

In 1914, evening observations from the day prior to the bulletin’s issue date were added too. In the following years, new observations were gradually introduced, primarily to provide a more comprehensive set of variables to meet the needs of aviation, which became the bulletin’s main scope in 1927.

3 ReData activities

3.1 Scanning the RCMO weather bulletins

The first achievement of the ReData project was the scanning of the daily RCMO weather bulletins which are now freely available as JPG files (https://doi.org/10.13130/RD_UNIMI/R1GVKF - [26]) from December 1879 to April 1940, totaling over 60 years of data and comprising more than 22,000 bulletins. As each bulletin was scanned page by page, the collection amounts to 99,518 individual scans, with formats ranging from 23 × 31 cm for the oldest bulletins to 35 × 50 cm for the most recent ones. Most of the scans (about 95%) were made from bulletins stored at the Department of Environmental Science and Policy of UniMi, while the remaining scans were obtained from bulletins preserved at the Central Historical Library of the Collegio Romano in Rome.

The scans of the bulletins from UniMi were carried out between 2016 and 2025 by the Garon private company; the remaining scans were completed in February 2025 as part of the “Dieci e Lode” project [27], funded by the Italian Ministry of Culture (<https://aisam.eu/progetti/10-e-lode-en/>, accessed on 1st July, 2025). To scan all sheet documents with dimensions equal to, or smaller than, A3 format, we used a Xerox flatbed scanner at a resolution of 300 dpi, calibrated to the original document size, while for documents larger than A3, we employed a portable L-shaped scanner, the Czur Shine Ultra Pro, equipped with a 24-megapixel sensor.

Bollettino Meteorico
DELL'UFFICIO CENTRALE DI METEOROLOGIA
ROMA

N 365
31 Dicembre

Ore 8 antimeridiane

STAZIONI	Barometro a 0 ed al mare		Termometro centigrado		Vento	Cielo	Direzione delle nubi	Mare	Piegata in millimetri 24 ore	Temperature		OSSERVAZIONI DIVERSE	
	Millimetri	Diff. in 24 ore	Gradi	Diff. in 24 ore						Massima	Minima		
Belluno	—	—	—	—	—	—	—	—	—	—	—	—	
Donnadossola	768,4	+1,9	11,4	-2,6	calma	1/2 copo	—	—	—	10,9	-0,5	Brina nella notte	
Milano	768,1	+1,2	11,4	-1,9	W	3/4 cop	WNW	—	—	9,8	11,3	—	
Verona	767,4	+1,6	11,2	-0,8	E	0	nebbioso	—	—	9,6	3,3	—	
Venezia	767,9	+1,6	5,4	-0,2	WNW	9	nebb.	calmo	—	7,9	11,1	Nebbie fra le 14 ore.	
Torino	768,4	+1,2	0,4	-1,0	SW	9	nebb.	—	—	6,6	0,2	—	
Parma	768,0	+1,9	11,1	-2,6	SW	5	3/4 cop	—	—	11,1	—	—	
Modena	768,0	+1,9	5,8	-0,9	SW	19	nebbioso	—	—	11,9	5,1	Part. mag. fra le 14 ore.	
Genova	768,2	+1,9	12,0	+0,4	E	7	cop.	SE	molto	11,1	11,9	—	
Pesaro	—	—	11,2	-0,2	SW	8	1/2 cop.	—	calmo	—	11,1	5,6	—
Porto Maurizio	768,1	+1,4	11,6	+0,6	NE	9	3/4 cop.	NE	molto	13,4	8,5	—	
Firenze	768,8	+1,7	10,0	+1,0	SW	debb.	cop.	SW	—	19,0	9,8	—	
Urbino	768,6	+0,8	9,0	+0,3	SW	2	sereno	—	—	11,0	9,0	—	
Ancona	768,4	+1,4	10,3	-1,0	W	debb.	sereno	—	calmo	—	11,1	9,8	—
Livorno	769,0	+2,5	13,1	+5,4	W	6	cop.	W	leg. mosso	—	19,3	12,3	—
Perugia	740,0	+1,9	7,9	+0,9	SSE	7	cop.	W	—	10,4	6,8	—	
Camerino	769,7	—	7,4	—	SSW	0	1/2 cop	WNW	—	9,9	6,8	—	
Portoferraio	769,8	+1,8	12,2	+0,7	calma	—	sereno	—	calmo	—	19,5	8,9	—

Fig 3. First page of the RCMO weather bulletin of 31 December 1882.

<https://doi.org/10.1371/journal.pclm.0000865.g003>

3.2 First attempt to digitize data

A first attempt to digitize the data in the RCMO weather bulletins was carried out using a web application developed by MNW, where users were entirely free to settle what to digitize, following a vertical approach, that is, once a city was selected, they could choose a time period spanning multiple available years, with each daily digitization corresponding to the completion of several data rows.

The idea behind this method was to encourage engaged citizen scientists to adopt a specific monitoring station, enabling them to consistently work on the same location, and thereby contribute to building long-term records of various meteorological variables reported in the bulletins. The results obtained using the MNW application were initially very encouraging: in fact, from the platform's launch in 2017, a significant interest was observed, with the number of rows entered per day peaking at nearly 300. Data entry remained relatively stable until July 2019, with spikes corresponding to the addition of new scans to the platform. Activity began to decline thereafter, with a temporary resurgence during the COVID-19 lockdown period. However, after September 2020, contributions significantly dropped. Over the entire lifespan of the platform (which remains available, but no longer updated or actively maintained), a total of 21,287 observation rows were digitized, covering 39 cities and involving 88 users. The distribution of entries by city shows that, in the cases of Verona and Milan, user participation enabled the digitization of a substantial volume of data for the period 1879–1909. In particular, Verona saw the near-complete digitization of all available information, with a total of 7,520 rows entered, while Milan reached 4,590 digitized daily records.

From this experience, we derived three key take-home messages: (i) without the ability to engage a large number of potential collaborators, it is not feasible to digitize such a vast dataset; (ii) the vertical approach adopted was not the most effective way to exploit this type of dataset; and (iii) it is unrealistic to expect digitizing everything at once; therefore, it is essential to establish a prioritization strategy that focuses first on what is really possible to digitize. Based on these considerations, we opted for a new design of the digitization activity.

3.3 Setting up the data digitization on the Zooniverse platform

Zooniverse is an international platform that hosts projects across a range of disciplines, including astronomy, biology, and climatology. It provides volunteers with simple explanations, tutorials, and forums to guide their participation. Since its launch in 2009, it has engaged over 2.8 million volunteers, completing more than 900 million classifications across 95 active projects (<https://www.zooniverse.org/projects>, accessed on 1st January 2026).

Indeed, Zooniverse, which has already leveraged citizen science to successfully support historical meteorological and hydrological data recovery initiatives, such the monthly rainfall observations taken in UK [12] and SIREN (Saving Italian hydrological Measurements) [28] projects respectively, offers a valuable opportunity for ReData. The main feature of the projects on the Zooniverse platform is that they do not require specialized scientific knowledge, but involve processing large datasets and tasks that would be infeasible for a small group of researchers to handle individually.

Building on what we learnt from the digitization of the RCMO weather bulletins on the MNW platform, we decided to move our activity to the Zooniverse platform in 2024 and to design the digitization in order to (i) focus on pages containing the data of Italian observatories only, (ii) consider, at least in a first moment, a subset of the observatories (about 40) that can provide a good description of the spatial coverage of the reported variables over the Italian territory, (iii) adopt an horizontal approach focusing on one year at time which gives priority to produce spatial fields rather than single station long-term records, (iv) limit the digitization to the period 1882–1939, as prior to 1882 the bulletin underwent too many changes, and in 1940 bulletins were available only for the first four months and were issued in a reduced form. Overall, the total number of considered years is 58 including 21,183 bulletins. Since bulletins are scanned page by page, the amount corresponds to over 96,228 scans. Considering the number of meteorological variables recorded, the project estimates over 25 million individual data points that could be digitized through ReData.

Setting up the digitization project, the first task was to associate each scanned page with the corresponding observatories reported. Since the list of stations on each page remained consistent within the bulletins of a given year, these associations were established based on the first day of the year only. Accordingly, for each year, all the first pages of the daily bulletins were linked to a list of stations, all the second pages to another list, and so on. In addition, as Zooniverse does not accept files larger than 1 MB, it was necessary to slightly reduce the quality of any files exceeding this size, without affecting the document readability. An alternative option was resizing the images by cropping the edges, but this method

was discarded, because the scans are often skewed, and it was not feasible to individually check each scan, with a real risk of cutting off useful content.

The second task was defining the digitization workflows. For each year, a list of workflows was created, matching the number of municipalities cataloged in the bulletins. Each workflow is named after a specific city, and it groups all sheets containing information for that city. This structure allows for the digitization of all station-related data for a given year.

The third task involved defining the set of stations to be included in the digitization activities. We decided to consider approximately 40 cities listed in the bulletin up to 1893. For the subsequent period, although data availability in the bulletin significantly increases, we are going to maintain the same number of stations, adding new ones only to continue the records of stations that ceased to be reported or to cover areas that were previously unrepresented (e.g., stations in former Italian colonies).

The fourth task regards the variables to be included in the digitization activities. We adopted the same approach used for station selection: all variables reported up to 1913 were digitized, and, where possible, the same set of variables will be selected for the subsequent years.

In setting up the digitization project, we chose to digitize all data from scratch, even though a small portion had already been digitized using the MNW application. This decision was made because the previously digitized portion was relatively small and, in that earlier effort, each data entry was digitized only once, whereas the new digitization protocol requires each entry to be digitized three times [9]. Nevertheless, the dataset produced by MNW will still be useful for cross-checking the newly digitized data.

3.4 Ethics Statement

This publication uses data generated via the Zooniverse.org platform under the Zooniverse User Agreement and Privacy Policy (<https://www.zooniverse.org/privacy>). The authors do not have access to any personally identifiable information about the volunteers.

3.5 ReData on Zooniverse

The current section outlines the structure of the ReData project on the Zooniverse platform that enables users contributing to data digitization in a simple and intuitive way, without requiring explicit technical or scientific skills. The homepage (Fig 4) includes a navigation menu, a blue banner for displaying announcements, a brief overview of the project, a list of active workflows, and user classification statistics. From the menu, users can access the following sections: (i) more information about the project, including details on the research team, achieved results, and frequently asked questions (FAQs) (via *About* tab); (ii) a randomly selected classification task from the active workflows (via *Classify*); (iii) a discussion forum (via *Talk*) where users can interact with other volunteers and the research team; (iv) the broader Zooniverse collections (via *Collect*); and (v) the user's recent classification activity (via *Recents*).

The *Talk* section is particularly important for the project's success, as it allows volunteers to share questions or suggestions about bulletins and workflows, and to report potential errors or inconsistencies. This continuous interaction has enabled us to refine the data collection process and ensure higher data quality.

In the central part of the homepage, users can select a specific workflow, identified by station name and year. However, the individual classification task (i.e., the specific day to be digitized) is randomly selected from the available entries within that workflow. Once a workflow is selected, either from the central page or the *Classify* menu, a scanned image from the bulletin appears to the volunteers. Then, they complete a series of tasks to digitize all variables for the specified station, year, and day, thereby completing the classification.

To accommodate user preferences, all workflows for a given year are made available simultaneously, allowing volunteers to choose based on their interests, whether a coastal city, a major urban area, or a specific region. As previously mentioned, we upload one year at a time. This approach allows us to showcase project progress gradually, which we have found to be more

ANNOUNCEMENT: If you feel like it, fill the [Google Form](#) with the country where you are from/where you are classifying from!
ANNOUNCEMENT (17 MAY 2025): Thanks to your help, we completely classified 1885! See the [related announcement section](#) for more details!

ABOUT CLASSIFY TALK COLLECT RECENTS

ReData

Help us digitize Italian meteorological data from 1879 to 1940 to better understand the climate and its past changes!

Learn more

STAZIONI

ROMA

OSSERVAZIONI DIVERSE

ABOUT REDATA

ReData (acronym of Recovery of Data) is the project that the *Meteonetwork Association* is carrying out together with the *University of Milan* and *Politecnico di Milano*. The project consists of the digitalisation of all meteorological data collected from 1879 to 1940 which were archived at the Royal Central Meteorological Office.

The preservation of these data, that risks being lost or deteriorated over time, is crucial.

The aim of ReData is to create a public dataset available to anyone who wants to carry out research in the climatological or historical field.

1.311	13.888	12.775	321
Volunteers	Classifications	Subjects	Completed Subjects

Get started

The bulletin contains data collected from several meteorological stations, each workflow is related to one of them. Choose your favorite! Now we are digitizing the year 1886!

Foggia 1886	Roma 1886	Aquila 1886	Chieti 1886	Verona 1886	Venezia 1886
Torino 1886	Alessandria 1886	Parma 1886	Modena 1886	Genova 1886	Forlì 1886
Pesaro 1886	Porto Maurizio 1886	Firenze 1886	Ancona 1886	Livorno 1886	
Perugia 1886	Camerino 1886	Portoferraio 1886	Bari 1886	Napoli 1886	
Portoferraio 1886	Potenza 1886	Lecce 1886	Cosenza 1886	Cagliari 1886	
Cozzo Spadaro 1886	Catanzaro 1886	Reggio di Calabria 1886	Palermo 1886	Catania 1886	
Caltanissetta 1886	Portompedocce 1886	Siracusa 1886			

Fig 4. Overview of the Re-Data project homepage on Zooniverse (<https://www.zooniverse.org/projects/meteonetwork/redata>, accessed on 1st January 2026) with the main features of the platform.

<https://doi.org/10.1371/journal.pclm.0000865.g004>

motivating for volunteers than waiting until the entire project is completed. Additionally, it enables us to download and inspect data incrementally, helping to monitor and ensure the quality of the digitized information. The required sequence of tasks (Fig 5), which allows a classification to be completed, follows the layout of the observation table, proceeding from left to right across a row.

Every task includes a description of variables to be digitized, a data entry field (either a text box or a dropdown menu), and a link to help documentation to assist users. The tasks are seven. The first allows us to check whether the selected station is actually present in the scanned image, the second to add a line in the image of the scanned page to avoid confusing rows in the tables. The remaining five allow to insert the following information:

- sea level pressure (SLP) and corresponding 24 h tendency;
- temperature and corresponding 24 h tendency;
- wind direction and speed;
- sky condition, clouds direction and sea state;
- maximum and minimum temperature and total precipitation in the last 24 hours.

1 TASK TUTORIAL
Can you find Firenze in the first column?
Yes
No
Yes, but the row is all blank
NEED SOME HELP WITH THIS TASK?
Next →

2 TASK TUTORIAL
(Firenze) If you're working on a row far from the table header, adding a highlighted line could be very helpful. It's essential to avoid confusing the rows in the table.
Draw a line to highlight the row where you are working
NEED SOME HELP WITH THIS TASK?
Back Next →

3 TASK TUTORIAL
The second column reports the barometric pressure (Millimetri).
Write below the corresponding data.
762.9
NEED SOME HELP WITH THIS TASK?
The third column reports the difference in terms of barometric pressure in 24h (Diff. za in 24 ore).
Write below the corresponding value with its sign.
1.2
NEED SOME HELP WITH THIS TASK?
Back Next →

4 TASK TUTORIAL
The fourth column reports the temperature (°C) (Gradi).
Write below the corresponding data with its sign.
17
NEED SOME HELP WITH THIS TASK?
The fifth column reports the difference in terms of temperature in 24h (Diff. za in 24 ore).
Write below the corresponding value with its sign.
1
NEED SOME HELP WITH THIS TASK?
Back Next →

5 TASK TUTORIAL
The sixth column reports the wind direction (Direzione).
From the list below, select the wind direction code related to the city you are considering.
Wind direction
NW
NEED SOME HELP WITH THIS TASK?
The seventh column reports the hourly wind velocity (or the estimation of the wind force) (Velo. oraria o velocità stimata). If the data is present, write it below.
2
NEED SOME HELP WITH THIS TASK?
Sometimes, in this field you can find a short description instead of the value. If this is not the case, left the box blank and it will count as NA.
Wind force
NA
NEED SOME HELP WITH THIS TASK?
Back Next →

6 TASK TUTORIAL
The eighth column reports the sky condition (Cielo).
Select the correct option in the list below.
Sky conditions
sereno
NEED SOME HELP WITH THIS TASK?
The ninth column reports the information about the clouds direction (Direzione delle nuvole).
From the list below, select the correct direction.
Clouds direction
NA
NEED SOME HELP WITH THIS TASK?
The column n.10 reports the sea condition (Mare).
From the list below, select the correct condition.
Sea conditions
NA
NEED SOME HELP WITH THIS TASK?
Back Next →

7 TASK TUTORIAL
The column n.11 reports the maximum daily temperature (Massima).
Write below the corresponding value with its sign.
27.5
NEED SOME HELP WITH THIS TASK?
The column n.12 reports the minimum daily temperature (Minima).
Write below the corresponding value with its sign.
13
NEED SOME HELP WITH THIS TASK?
The column n.13 reports the millimeters of rain in the last 24h (Pioggia in mm in 24 ore).
If the data is present, write it below.
NA
NEED SOME HELP WITH THIS TASK?
Back Done & Talk Done

Fig 5. List of the tasks that are necessary to fill in in order to complete the classification.

<https://doi.org/10.1371/journal.pclm.0000865.g005>

The organization of these tasks has slightly changed from when we launched the ReData project on Zooniverse thanks to the engaged users' suggestions which helped us to optimize it. Moreover, moving on the right of the page, during the digitization it is possible to open a field guide (selecting *Field Guide*) where volunteers can find information about the history of the RCMO weather bulletins, an explanation on how to interpret some symbols during the digitization and the main FAQs. In particular, the *Field Guide* focuses on providing in-depth knowledge about the bulletins, including insights and curiosities related to the project. While optional, it enhances volunteer engagement by broadening their understanding of the project's significance.

Another essential section of the platform includes the tutorial, which is available during all steps of the digitization (selecting *Tutorial*). It provides a brief walkthrough of the project's goals and tasks, guiding volunteers on how to perform classifications.

3.6 Beta review process

Before the official launch, the project underwent a two-stage review process. The first stage, an internal review, involved a thorough evaluation by Zooniverse team members. During this phase, the project structure and page content were assessed to ensure they met specific standards, and detailed feedback was provided, including suggestions for potential

improvements. The second stage was the beta testing phase, where the project was introduced to a limited group of Zooniverse volunteers, which participated by submitting beta classifications and sharing their comments via a standardized Google form.

These phases were critical as they allowed for ultimate adjustments before the final launch on the platform, and they also tested data aggregation and analysis methods designed for the project. For beta testing, a reduced subject set was required to effectively assess the workflows during this period. Since these workflows planned for the ReData project were identical, only differing by station they referred to, a single workflow was proposed for the beta phase (Belluno 1882). A reduced subject set of 181 daily scans, created from January to June 1882 (six months in total), was associated with it. This testing phase provided an opportunity to gauge the responsiveness of the Zooniverse community: over 900 classifications were completed in less than two weeks. In this initial phase of the project, the discussion board proved to be highly valuable, as volunteers shared questions and observations that were instrumental in improving the site. The project was officially launched on 26th November, 2024.

3.7 Key challenges

Although the primary activity of the project is straightforward and schematic, and volunteers are simply required to carefully copy data recorded in the bulletins, there are some intrinsic challenges presented by documents, paramount to this project.

Firstly, handwriting poses significant encounters. Once users select a workflow, they must verify that the station (or city) name, written in an old italic style, appears in the first column of the bulletin. Then, users must transcribe a large number of values. At first glance, this may appear simple, but it is not always clear-cut. For instance, the same digit is sometimes written in different styles, which may introduce uncertainty. The primary challenges identified so far involve the numbers 5 and 7. Many of these ambiguities have been solved by cross-referencing the second and fourth columns of the bulletin. These columns contain information on the 24-hour differences in pressure and temperature, which link data of a specific day to that of the previous day; through such cross-checking, it has been possible to accurately decipher unclear handwriting.

In certain tasks, such as identifying wind and cloud directions, users are asked to recognize an acronym in uppercase letters representing one of the 16 cardinal points. Since the response options are limited in these cases, a dropdown menu is added to facilitate the process. In other tasks, users must recognize words, often adjectives describing sky, sea, and wind. Many of these terms are frequently repetitive, thus, a dropdown menu is implemented as well. However, for non-Italian-speakers, decoding certain words can be difficult, particularly when abbreviations are used (e.g., *ab. forte*, *deb.*). Some examples are provided in [Fig 6](#).

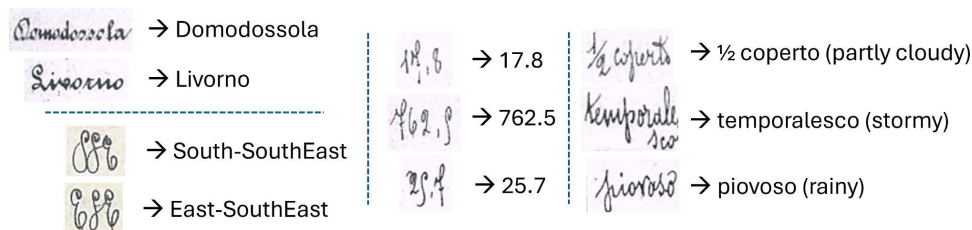


Fig 6. Examples of how to decipher calligraphy for numbers, words and acronyms taken from the Field Guide, which volunteers can access at all times.

<https://doi.org/10.1371/journal.pclm.0000865.g006>

3.8 The first five months on Zooniverse: project progress and community engagement

Six months after its launch, the digitization of the first four years (1882–1885) was completed at a rate of approximately one year every five-six weeks (data updated to 15th May, 2025). Each owner of the project can directly download data from the Zooniverse page. These are included in one file saved in CSV format, where each row represents a classification (i.e., a day for one station with all available variables), and each column corresponds to a task saved as a string in JSON format. In addition to the digitized data, a substantial number of metadata is available for all classifications, allowing to monitor the progress of the project while preserving the anonymity of the volunteers.

Specifically, we digitized 34 stations for the year 1882 and 40 stations for the subsequent years, resulting in a total of 171,437 classifications. At first, we checked the data to ascertain whether all the expected days for each station and year were included in the dataset, verifying the correct working of the platform. For our project, we required three classifications per day; however, we observed that if a workflow is not retired immediately upon completion, it remains available and may be classified again. Therefore, careful monitoring of classification progress is necessary to ensure workflows are promptly retired; otherwise, it sometimes occurs that volunteers classify the same day multiple times.

By the end of this analyzed period, thanks to the internationality of the Zooniverse platform, 1,153 volunteers completed at least one classification, with participation increasing almost linearly. After an initial peak of 303 volunteers, the number of active participants per week stabilized between 50 and 100 (Fig 8A), ensuring a weekly mean of 4,762 classifications. The number of daily classifications (Fig 8B) was highly variable and, as expected, followed a weekly cycle, with higher values during weekends and holidays (e.g., Easter).

An analysis of the classifications grouped by username shows that 47.8% of all classifications were completed by three volunteers, and 29.4% were performed by a single “Top Classifier” who systematically catalogued all available data. When new workflows are activated, the Top Classifier continuously works until completing all classifications, explaining some of the high peaks in daily activity. Considering the start and end times for each classification, a median classification time of approximately one minute is found (Fig 8C). While this time varies from month to month, it gives evidence of a decrease after March due to improvements in the workflow structure that enhanced efficiency.

In particular, 50% of classifications were completed within 30–100 seconds (Fig 8D), 24% took longer than 100 seconds, and 26% took less than 30 seconds, with 22.4% of classifications being completed within 10–30 seconds. Longer classification times may reflect volunteers reading the tutorial during classification or leaving the computer unattended, whereas shorter times often occur when volunteers do not find the required town or encounter an empty line. It is estimated that the minimum time required to scroll through all tasks is about 12 seconds. Finally, it is noteworthy that most classifications (95.7%) completed in 10–20 seconds were performed by the Top Classifier, whose experience allows for significantly faster processing.

Interestingly, the data also allow us to infer the UTC time zone of each volunteer (Fig 9A). Their spatial distribution indicates that most volunteers reside in the USA and Europe. Some inaccuracies may exist, as Zooniverse does not account for the transition between summer and winter time. Furthermore, considering only the subset of 188 volunteers (16.3% of the total) who declared their country of origin (Fig 9B), 39% were from the USA, 10% from the UK, and 8%, 7%, and 4% from Italy, Germany, and France, respectively. The remaining volunteers came from 37 other countries.

3.9 An overview of the first four digitized years (1882–1885)

Thanks to having three digitizations for each day, we extracted all values using an automated procedure prepared by us and independent from the Zooniverse platform. First, all digitized information (55,520 values for each variable) was processed by treating as missing any character string where a numeric value was expected, as well as any empty fields where a value should have been selected from a dropdown menu. Additionally, for precipitation data, both missing values and the absence of precipitation are indicated in the bulletins by a solid line. Consequently, all corresponding digitizations

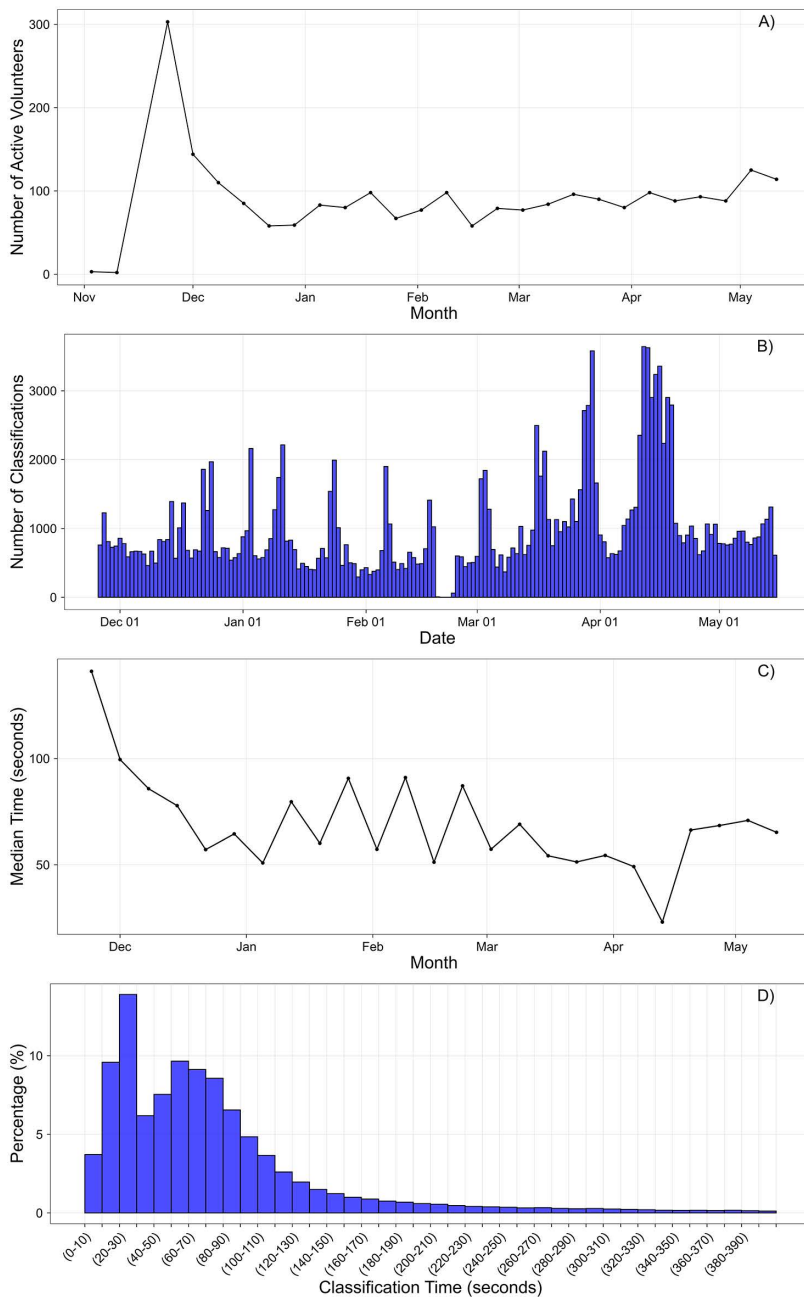


Fig 8. A) Weekly number of active volunteers; B) Daily number of completed classifications; C) Weekly median time required to complete a classification; D) Percentage of classifications for each time interval. The information refers to the period between 26th November, 2024 and 15th May, 2025.

<https://doi.org/10.1371/journal.pclm.0000865.g008>

recorded as 'not available' (NA), 0, blank entries, or other symbols were interpreted as 0 when both maximum and minimum temperature values were available; otherwise, they were treated as missing.

Next, we checked and extracted the values where at least two digitizations were equal. The results showed that this was the case for 98.6 to 99.8% of the data, depending on the variable considered (Table 1). The lowest agreement was

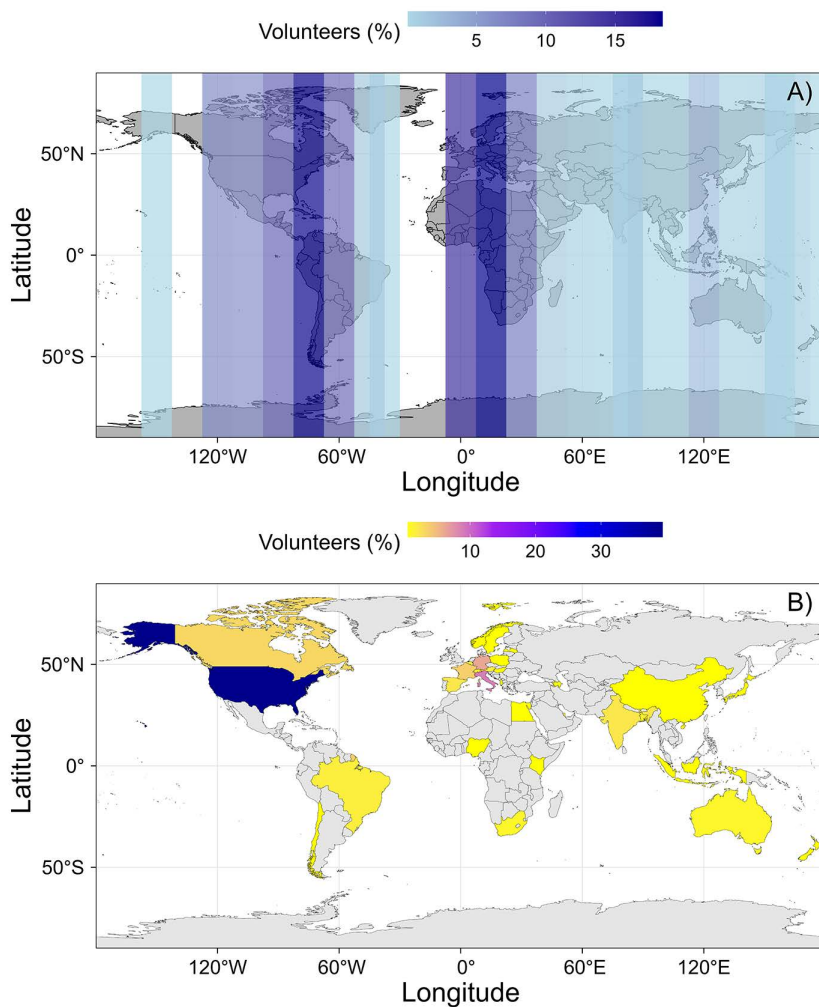


Fig 9. A) Percentage of volunteers for each UTC band; B) Percentage of volunteers for each country. The information about the borders comes from <https://www.naturalearthdata.com/>.

<https://doi.org/10.1371/journal.pclm.0000865.g009>

observed for wind speed. In most cases, the wind speed is provided as a numerical value, but some volunteers also selected a text option from the dropdown menu, set up only for a minority of classification where both the possibilities (number and text) were available, leading to a lower number of classifications with two identical values. For all other variables, this percentage exceeds 99.0%. In cases where all three values differed, the entries were stored in a separate file so that they could be reviewed in future to determine the correct value, for example by using information from other variables reported in the bulletin or from the preceding and subsequent days. For example, for temperature and pressure, we reconstructed missing or invalid data from the previous day's values and the 24-hour tendency, or vice versa. After this integration, the data availability for these four variables reached 99.9%.

Considering that the mean number of classifications per week is 4,762, the estimated time required to complete the digitization of all data is approximately nine years. For this reason, we evaluated the percentage of data with two identical digitizations, assuming only two classifications per entry (Table 1), investigating in this way the possibility to reduce the amount of digitization from three to two. The results show that, even with only two digitizations, the percentage of

Table 1. Percentage of matching values when considering three classifications and percentage of matching values when considering only two classifications. The results are reported for all variables considered, with parentheses indicating whether the information comes from a drop-down menu. Percentages of matching classifications are also reported after integration with information from other variables. The symbol * is used for variables for which the integration of information from other variables has not been applied.

Variable	Three classifications		Two classifications	
	N° of two equal values	N° of two equal values (integrated)	N° of two equal values	N° of two equal values (integrated)
Pressure	99.6%	99.9%	93.6%	97.2%
Barometric trend	99.7%	99.9%	95.4%	97.5%
Temperature	99.8%	99.9%	95.8%	97.6%
Thermometric trend	99.7%	99.9%	95.7%	97.6%
Maximum temperature	99.6%	*	94.6%	*
Minimum temperature	99.7%	*	94.7%	*
Rainfall	99.8%	*	96.7%	*
Sky (dropdown)	99.6%	*	96.8%	*
Cloud direction (dropdown)	99.1%	*	98.1%	*
Sea condition (dropdown)	99.4%	*	98.4%	*
Wind direction (dropdown)	99.8%	*	96.8%	*
Wind direction	99.0%	*	94.7%	*
Wind speed (dropdown)	98.6%	*	98.3%	*
Wind speed	99.6%	*	97.3%	*

<https://doi.org/10.1371/journal.pclm.0000865.t001>

extracted data ranges from 93.6% to 98.4%. This percentage increases significantly for temperature and pressure when they are integrated using information from the previous day and the 24-hour tendency with values between 97.2% and 97.6%. Therefore, considering the good results and the important advantage in choosing two digitizations instead of three in terms of time, we have decided to continue the project with two digitizations, reducing the expected completion time to six years.

4 Reconstruction of the flood case event

A noteworthy benefit by the ReData project lies in its contribution to improving the insight of past atmosphere circulation over southern Europe, both through the direct information provided by the bulletins and the contribution that these data can give to reanalysis datasets such as the NOAA 20CR which are crucial for depicting a comprehensive picture of climate evolution over time as well. These datasets enable the reconstruction of historical synoptic conditions, allowing for detailed analysis of specific meteorological events, such as storms, floods, and droughts [29–32]. Moreover, reanalysis datasets facilitate the reconstruction of meteorological conditions in remote areas lacking observational data. However, the accuracy of these models significantly improves with a denser network of observational stations across the region.

Studies by [33] and [34] highlighted the number of European stations that contribute to reanalysis datasets. Examining the historical period covered by the RCMO bulletins reveals that stations on the Italian peninsula are sparse (Fig 10), with no more than a dozen in total. This suggests that making the data recovered from these bulletins available to the scientific community strongly helps filling the observational gap in Italy, thereby enhancing the accuracy of reanalysis datasets.

To better highlight the contribution of our RCMO weather bulletins, one of the most catastrophic hydro-meteorological events of the 19th century [35], occurring in September 1882, is illustrated as a showcase. This event was characterized by a long period of heavy rainfall in northern Italy with maximum values recorded between September 15th and September 17th. The peak of precipitation in this 3-day period was recorded between Vicenza and Trento cities (in particular, 596 mm in Posina, 417 mm in San Ulderico del Tretto, 407 mm in Valle dei Signori, 375 mm in Schio); nevertheless, the area

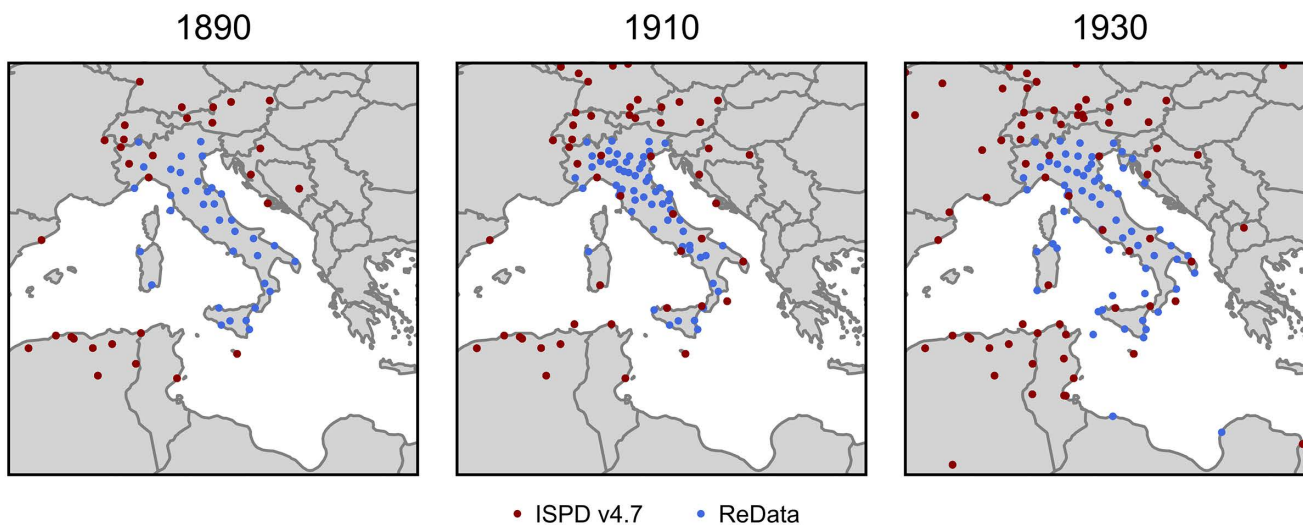


Fig 10. Distribution of SLP data available with the International Surface Pressure Databank (ISPD version 4.7, in maroon) and ReData (in blue) stations in the year 1890 (left), 1910 (center) and 1930 (right). The information about the borders comes from <https://www.naturalearthdata.com/>.

<https://doi.org/10.1371/journal.pclm.0000865.g010>

experiencing more than 200 mm was very large covering a great fraction of the eastern part of the southern Alps [36]. From a hydrological perspective, this intense rainfall caused widespread impacts across several regions, with the Adige River basin experiencing the most severe damage. Multiple areas, including Trento and Verona, were affected by significant flooding. In Verona, the Ponte Nuovo bridge collapsed on September 14th, and further river overflows occurred on September 17th (Fig 11), resulting in extensive destruction and several dozen fatalities. Downstream, at dawn of September 18th, a 290-meter breach at Legnago city caused massive flooding in the Valli Grandi Veronesi and Polesine areas, situated between the Adige and Po Rivers, resulting in considerable devastation. The flood caused significant material destruction across Italian territory: 17,000 hectares were submerged on the left bank and 109,000 hectares on the right. A total of 62 municipalities and 138,000 inhabitants were affected. Approximately 540 homes were destroyed and 8,200 were damaged. Forty major bridges were washed away, and 2,500 hydraulic structures, including small bridges, supports, drains, and manholes, collapsed [37]. This unprecedented episode transformed the regional disaster into a matter of national concern, particularly on flood management strategies, environmental policies, and urban planning [38].



Fig 11. Archive images of Verona flooding in 1882. A) City map where flooded areas are darker (source from https://it.wikipedia.org/wiki/Inondazione_di_Verona_del_1882, accessed on 1st July, 2025); B) Ponte Nuovo's collapse (source from https://it.wikipedia.org/wiki/Alluvione_del_Polesine_del_17_settembre_1882, accessed on 1st July, 2025); C) plate marking the level reached by the water on 17th September 1882 in Verona city..

<https://doi.org/10.1371/journal.pclm.0000865.g011>

The bulletins' synoptic charts of this 3-day period (Fig 12B, D, F) provide valuable information about the event, which cannot be spotted by the corresponding charts from the NOAA 20CR (Fig 12A, C, E). The day on which the Twentieth Century Reanalysis shows the most pronounced differences compared with the chart and the data from the RCMO bulletins is September 15th, 1882. The very clear low pressure center on the Gulf of Genua (about 750 millimeters of mercury, equal to about 1000 hPa) evident from Fig 12A is in fact very poorly captured by the 20CR (Fig 12B) and also the projection of the gridded reanalysis values onto the Italian stations (Table 2) gives evidence of sea level pressure values strongly overestimated (e.g., +8.9 hPa for Parma, +8.8 hPa for Genua, +8.5 hPa for Milan). On the other hand, the reanalysis of sea level pressure values slightly underestimates the stations' ones in the eastern

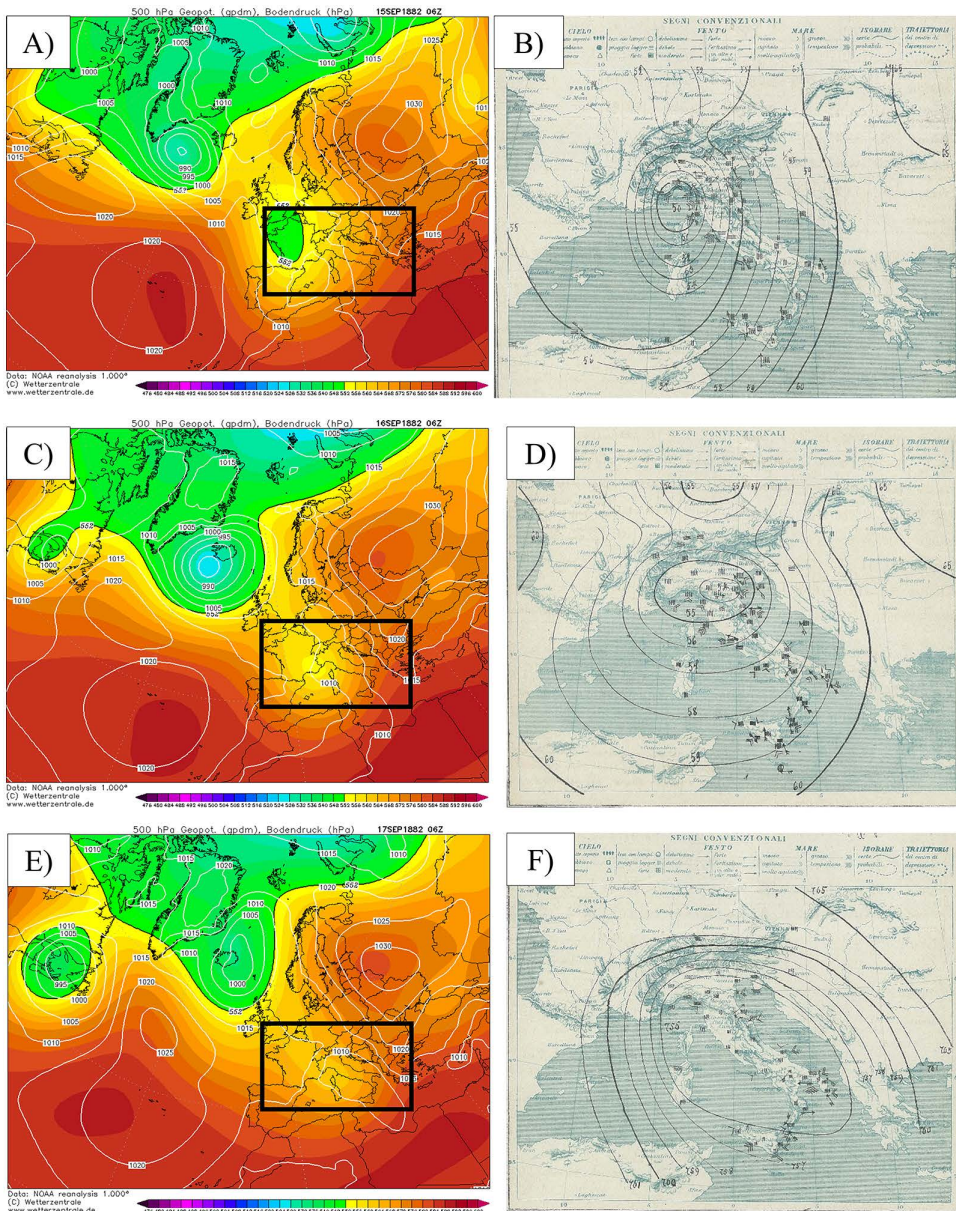


Fig 12. Comparison between reanalysis (left) and hand-drawn (right) synoptic maps on September 15th (A, B), 16th (C, D) and 17th (E, F), 1882.

<https://doi.org/10.1371/journal.pclm.0000865.g012>

Table 2. Comparison of sea-level pressure (SLP) values reported in RCMO bulletins at selected Italian stations with data from the 20CR on September 15th, 1882.

ID	Weather station	Lat	Lon	RCMO bulletin data		NOAA 20CR	Delta [hPa]
				SLP [mm Hg]	SLP [hPa]	SLP [hPa]	
1	Domodossola	46.11	8.29	754.4	1005.8	1008.7	-2.9
2	Milano	45.46	9.19	751.1	1001.4	1009.9	-8.5
3	Verona	45.44	10.99	752.6	1003.4	1010.4	-7.0
4	Venezia	45.44	12.32	755.6	1007.4	1011.3	-3.9
5	Torino	45.07	7.69	751.8	1002.3	1008.8	-6.5
6	Parma	44.80	10.33	750.6	1000.7	1009.6	-8.9
7	Modena	44.65	10.93	751.5	1001.9	1009.6	-7.7
8	Genova	44.41	8.95	749.8	999.7	1008.5	-8.8
9	Pesaro	43.91	12.91	754.4	1005.8	1010.1	-4.3
10	Porto Maurizio	43.89	8.03	749.4	999.1	1007.0	-7.9
11	Firenze	43.77	11.26	752.1	1002.7	1009.3	-6.6
12	Urbino	43.73	12.64	755.1	1006.7	1010.0	-3.3
13	Ancona	43.62	13.52	755.3	1007.0	1010.4	-3.5
14	Livorno	43.55	10.31	750.3	1000.3	1008.1	-7.8
15	Perugia	43.11	12.39	755.1	1006.7	1010.0	-3.3
16	Camerino	43.13	13.07	754.2	1005.5	1010.4	-4.9
17	Porto Ferraio	42.81	10.32	751.2	1001.5	1007.4	-5.9
18	Aquila	42.35	13.40	755.5	1007.3	1010.9	-3.7
19	Roma	41.90	12.50	754.7	1006.2	1010.3	-4.1
20	Foggia	41.46	15.54	758.3	1011.0	1011.8	-0.9
21	Napoli	40.85	14.27	757.0	1009.3	1011.9	-2.7
22	Portotorres	40.83	8.40	750.4	1000.5	1005.6	-5.1
23	Potenza	40.64	15.81	759.2	1012.2	1012.4	-0.2
24	Lecce	40.35	18.17	760.3	1013.6	1012.9	0.8
25	Cosenza	39.30	16.25	759.9	1013.1	1012.8	0.3
26	Cagliari	39.22	9.12	754.0	1005.3	1007.2	-1.9
27	Catanzaro	38.91	16.59	758.9	1011.8	1012.9	-1.1
28	Reggio di Calabria	38.11	15.65	759.0	1011.9	1012.6	-0.6
29	Palermo	38.12	13.36	756.3	1008.3	1011.3	-2.9
30	Caltanissetta	37.49	14.06	760.0	1013.3	1012.1	1.1
32	Siracusa	37.08	15.29	760.3	1013.6	1012.5	1.1

<https://doi.org/10.1371/journal.pclm.0000865.t002>

part of Sicily and Apulia (e.g., -1.1 hPa for Siracusa, -0.8 hPa for Lecce), contributing to further bias in evaluating the pressure gradients over Italy. Due to these errors, the 15 hPa pressure gradient from southeastern Italy (1013.6 hPa recorded in Siracusa and Lecce) to the Gulf of Genua (999.1 hPa recorded in Porto Maurizio and 999.7 hPa recorded in Genua), turns out to be reduced to about 5 hPa only. The same problem is present also in the following two days even though the errors of the 20CR are lower. Specifically, for September 16th (Fig 12C, D) the 9 hPa pressure gradient from southeastern Italy (1014.3 hPa recorded in Lecce) to the minimum in the eastern part of the Po Plain (1005.3 hPa recorded in Parma and Modena) is reduced to 5.6 hPa, and for September 17th (Fig 12E, F) the 3.2 hPa pressure gradient from Sicily (1010.3 hPa as mean of the values recorded in Palermo, Caltanissetta and Siracusa) to the Po Plain (1007.1 hPa, mean of the values recorded in Turin, Milan, Verona, Venice, Parma and Modena) is completely

eliminated (0.1 hPa). It is therefore remarkable to highlight how the hand-drawn maps from the RCMO bulletins well identify the low-pressure centers on these three days, while the coarser NOAA reanalysis maps fail to accurately capture them. In the RCMO maps it is also evident the counterclockwise circulation which activated southeasterly winds (scirocco) along the Adriatic Sea that significantly increased temperatures and melted the snow fallen on the Alps in the previous weeks which, combined with heavy precipitation and saturated soils, led to major flooding and overflows in nearly all rivers in the Veneto and Trentino regions.

It is worth noticing that beside the significant errors in the sea level pressure fields over Italy, the 20CR charts give evidence of typical features of heavy precipitation events over northern Italy as the marked trough of the 500 hPa geopotential height over western Europe and the blocking anticyclone over eastern Europe. However, this case study clearly demonstrates the importance of recovering new data and integrating historical records into modern datasets to address gaps and improve the accuracy of reanalysis.

5 Discussion and conclusions

In recent decades, many countries around the world have launched climate data rescue projects to preserve meteorological information stored in fragile paper records and make it accessible to the scientific community. These initiatives aim to enhance the accuracy of climatological studies and historical reconstructions of weather events. In this context, global reanalysis datasets are particularly important, as their reliability depends on the consistency and spatial coverage of the underlying past observations.

Recovering historical observations enables the construction of long-term climate data series, which are essential for improving our understanding of the Earth system and for increasing the accuracy of both past and future climate analyses. Over the last two decades, numerous efforts have contributed to this goal [39], including major initiatives such as the Atmospheric Circulation Reconstruction over the Earth (ACRE) project [40,41].

This paper describes the methodological approach adopted and the activities undertaken within the ReData project, a venture promoted by the Meteonet network association and the University of Milan, which aims to digitize meteorological observations contained in the RCMO bulletins. Specifically, the core of this work lies in the development of the ReData digitization process, and it presents the methodology built on Zooniverse as a climate data rescue product, fully replicable by anyone (scientists, volunteers, etc.).

After carefully evaluating how to manage a large volume of scanned bulletins, the main challenge was to design an accessible digitization workflow that could actively engage volunteers, including those without a scientific background. In the months following the official launch, the project received significant contributions from volunteers worldwide, highlighting the importance of regular feedback and engagement with the research team. In particular, through daily exchanges with the most active volunteers, we have fostered a keen sense of community that prompted this project not only to facilitate the recovery of historical meteorological data, but also to create a global network of enthusiasts. To recognize their effort, we have also sent to the Top Classifiers a digitizer's certificate signed by the institutions that coordinate the project.

The overall project is based on the RCMO weather bulletins available from 1879 to 1940, covering nearly 61 years and including more than 25 million individual data points. Each daily bulletin contains at least 13 handwritten columns. Due to the complexity of the material, faint ink, varying handwriting styles and evolving formats, standard OCR or machine learning methods are ineffective [42,43]. The handwriting not only varies between years, but sometimes from day to day, likely due to contributions from multiple individuals. Earlier records from the 19th century are particularly challenging, as the characters and number forms often differ from modern conventions. In such cases, the human eye remains the only reliable tool, though it is more time-consuming.

This is where the Zooniverse platform plays a crucial role, since it enables crowdsourcing of this monumental task through a global network of dedicated citizen scientists. The project has maintained sustained interest, and thanks to volunteer support, over five years' worth of bulletins have already been digitized. While volunteers have been extraordinary,

the research team has also played a key role, setting up the Zooniverse infrastructure, maintaining communication with contributors, and resolving issues related to data interpretation and platform functionality; these can be considered useful recommendations for similar data rescue activities. Recognizing that volunteers prefer quick and engaging tasks, the team has worked to streamline the digitization process. We moved from a very single-task interface, where volunteers had to input data one-by-one, to a less fragmented view in which they can see and insert data more easily. This also had the effect of reducing the loading time between the different tasks, sensibly reducing the total time required for a classification. When looking at the median time required, this totals 58.2 seconds per classification. Keeping into account all cities involved and available days, the 1882–1885 period has required a total of 2700 man-hours, an amount that would have been unachievable without the use of a pool of volunteers such as the one offered by Zooniverse.

Currently, the project continues the digitization process in a structured and efficient manner. This effort not only ensures the long-term preservation of a valuable meteorological archive, but also enhances its accessibility to the scientific community, supporting more accurate historical climatological reconstructions. As more bulletins are digitized through the ReData project, thousands of data points will become available for analysis in the coming months and will be shared with the public. To demonstrate the scientific value of the digitized data, a case study was conducted, with the reconstruction of an extreme precipitation event that affected northeastern Italy in September 1882, a historically and culturally significant episode. This case study displays the potential of recovered data to improve the reconstruction of past atmospheric circulation patterns. The preservation and digitization of historical climate data will benefit a diverse audience, including students, researchers, institutions, businesses, and weather services, facilitating deeper insights into past meteorological events and ongoing climate change.

Acknowledgments

The authors would like to thank the Meteonetwork association for its continuous support in this project. A special mention goes to the Zooniverse volunteers who daily provide a massive help in digitizing meteorological data. Lastly, we thank the Garon company for its massive job in scanning all the RCMO bulletins during these last years.

This publication uses data generated via the [Zooniverse.org](https://www.zooniverse.org) platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

Author contributions

Conceptualization: Alessandro Ceppi, Veronica Manara, Marco Giazzi, Maurizio Maugeri.

Data curation: Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Luca Cerri, Maria Di Giovanni, Ludovico Lapo Luperi, Luca Ronca, Elisa Sogno.

Formal analysis: Veronica Manara, Gabriele Buccheri, Goffredo Caruso, Luca Cerri, Maria Di Giovanni, Ludovico Lapo Luperi, Elisa Sogno.

Investigation: Alessandro Ceppi, Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Luca Cerri, Ludovico Lapo Luperi, Luca Ronca, Elisa Sogno, Maurizio Maugeri.

Methodology: Alessandro Ceppi, Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Ludovico Lapo Luperi, Luca Ronca, Elisa Sogno, Maurizio Maugeri.

Project administration: Alessandro Ceppi, Marco Giazzi, Maurizio Maugeri.

Resources: Marco Giazzi, Luca Ronca.

Software: Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Luca Cerri, Marco Giazzi, Ludovico Lapo Luperi, Elisa Sogno.

Supervision: Alessandro Ceppi, Veronica Manara, Marco Giazzi, Ludovico Lapo Luperi, Elisa Sogno, Maurizio Maugeri.

Validation: Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Ludovico Lapo Luperi, Elisa Sogno.

Visualization: Alessandro Ceppi, Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Luca Cerri, Marco Giazzi, Ludovico Lapo Luperi, Elisa Sogno, Maurizio Maugeri.

Writing – original draft: Alessandro Ceppi, Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Ludovico Lapo Luperi, Elisa Sogno, Maurizio Maugeri.

Writing – review & editing: Alessandro Ceppi, Veronica Manara, Yuri Brugnara, Gabriele Buccheri, Goffredo Caruso, Ludovico Lapo Luperi, Elisa Sogno, Maurizio Maugeri.

References

1. IPCC 2023; Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC. 2023. pp. 33–115.
2. Brunetti M, Maugeri M, Monti F, Nanni T. Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *Intl Journal of Climatology*. 2006;26(3):345–81. <https://doi.org/10.1002/joc.1251>
3. Brunetti M, Lentini G, Maugeri M, Nanni T, Auer I, Böhm R, et al. Climate variability and change in the Greater Alpine Region over the last two centuries based on multi-variable analysis. *Intl Journal of Climatology*. 2009;29(15):2197–225. <https://doi.org/10.1002/joc.1857>
4. Cornes RC, van der Schrier G, van den Besselaar EJM, Jones PD. An ensemble version of the E-OBS temperature and precipitation data sets. *JGR Atmospheres*. 2018;123(17):9391–409. <https://doi.org/10.1029/2017jd028200>
5. Morice CP, Kennedy JJ, Rayner NA, Winn JP, Hogan E, Killick RE. An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*. 2021;126(3).
6. Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. 2020;146(730):1999–2049.
7. Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, et al. The twentieth century reanalysis project. Vol. 137, *Quarterly Journal of the Royal Meteorological Society*. 2011:1–28.
8. Buizza R, Poli P, Rixen M, Alonso-Balmaseda M, Bosilovich MG, Brönnimann S, et al. Advancing Global and Regional Reanalyses. In: *Bulletin of the American Meteorological Society*. 2018;99(8):ES139–44. <https://doi.org/10.1175/bams-d-17-0312.1>
9. WMO 2024. Guidelines on the Best Practices for Climate Data Rescue. No. 1182. Geneva: WMO; 2024.
10. Slivinski LC, Compo GP, Whitaker JS, Sardeshmukh PD, Giese BS, McCol C, et al. Towards a more reliable historical reanalysis: improvements for version 3 of the Twentieth Century Reanalysis system. *Quart J Royal Meteorol Soc*. 2019;145(724):2876–908. <https://doi.org/10.1002/qj.3598>
11. Thorne PW, Willett KM, Allan RJ, Bojinski S, Christy JR, Fox N, et al. Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science. *Bull Amer Meteor Soc*. 2011;92(11):ES40–7. <https://doi.org/10.1175/2011bams3124.1>
12. Hawkins E, Burt S, McCarthy M, Murphy C, Ross C, Baldock M, et al. Millions of historical monthly rainfall observations taken in the UK and Ireland rescued by citizen scientists. *Geoscience Data Journal*. 2022;10(2):246–61. <https://doi.org/10.1002/gdj3.157>
13. Lorrey AM, Pearce PR, Allan R, Wilkinson C, Woolley JM, Judd E. Meteorological data rescue: citizen science lessons learned from Southern Weather Discovery. *Patterns*. 2022;3(6).
14. Ryan C, Duffy C, Broderick C, Thorne PW, Curley M, Walsh S. Integrating data rescue into the classroom. *Bull Am Meteorol Soc*. 2018;99(9):1757–64.
15. Mateus C, Potito A, Curley M. Engaging secondary school students in climate data rescue through service-learning partnerships. *Weather*. 2021;76(4):113–8.
16. Noone S, D'Arcy C, Donegan S, Durkan W, Essel B, Healion K, et al. Investigating the potential for students to contribute to climate data rescue: Introducing the Climate Data Rescue Africa project (CliDaR-Africa). *Geosci Data J*. 2024;11(4):758–74.
17. Noone S, Brody A, Brown S, Cantwell N, Coleman M, Sarsfield Collins L, et al. Geo-locate project: a novel approach to resolving meteorological station location issues with the assistance of undergraduate students. *Geosci Commun*. 2019;2(2):157–71. <https://doi.org/10.5194/gc-2-157-2019>
18. Camuffo D, Bertolin C. The earliest temperature observations in the world: the Medici Network (1654–1670). *Climatic Change*. 2011;111(2):335–63. <https://doi.org/10.1007/s10584-011-0142-5>
19. Manara V, Brunetti M, Beltrano MC, Bertoldi G, Brugnara Y, Cat Berro D, et al. Engaging high school students in rescuing and digitizing data from historical observations in Italy: the citizen science project Cli-DaRe@School. *Bull Am Meteorol Soc*. 2025;106(3):E509–24.
20. Manara V, Arcuri B, Beltrano MC, Bertoldi G, Brugnara Y, Brunetti M, et al. Italian precipitation records for the period 1921–1950 from the Citizen Science project Cli-DaRe@School. Zenodo. 2025. Available from: <https://zenodo.org/records/15084062>
21. Manara V, Arcuri B, Brunetti M, Beltrano MC, Bertoldi G, Brugnara Y, et al. A new dataset of Italian precipitation records for the period 1921–1950 from the Cli-DaRe@School citizen science project. *Bull of Atmos Sci & Technol*. 2025;6(1). <https://doi.org/10.1007/s42865-025-00111-3>

22. Giazzi M, Peressutti G, Cerri L, Fumi M, Riva IF, Chini A, et al. Meteonetwork: an open crowdsourced weather data system. *Atmosphere*. 2022;13(6):928. <https://doi.org/10.3390/atmos13060928>
23. Strasser BJ, Baudry J, Mahr D, Sanchez G, Tancoigne E. “Citizen Science”? Rethinking science and public participation. 2019.
24. Rubio-Iglesias JM, Edovald T, Grew R, Kark T, Kideys AE, Peltola T. Citizen science and environmental protection agencies: engaging citizens to address key environmental challenges. *Frontiers in Climate*. 2020;2.
25. Wehn U, Ghahesifard M, Ceccaroni L, Joyce H, Ajates R, Woods S, et al. Impact assessment of citizen science: state of the art and guiding principles for a consolidated approach. *Sustain Sci*. 2021;16(5):1683–99. <https://doi.org/10.1007/s11625-021-00959-2>
26. Ceppi A, Manara V, Brugnara Y, Buccheri G, Caruso G, Cerri L. ReData (Recovery of Data): daily meteorological bulletins edited by the Italian Royal Central Meteorological Office from 1879 to 1940. UNIMI dataverse; 2025.
27. Ceppi A, Brunetti M, Baldi M, Beltrano MC, De Vecchis E, Leali F, et al. The Dieci e Lode project: recovery of meteorological observations relating to the former Italian colonies. In: *Proceedings of Science*, 2025. 371. Available from: <https://pos.sissa.it/>
28. Mazzoglio P, Bertola M, Listo T, Princivalle L, Lombardo L, Viglione A, et al. Who is saving our streamflow data? Exploring volunteer profiles and their engagement in the SIREN data rescue project. *PLoS One*. 2025;20(10):e0333091. <https://doi.org/10.1371/journal.pone.0333091> PMID: [41066342](https://pubmed.ncbi.nlm.nih.gov/41066342/)
29. Stucki P, Rickli R, Brönnimann S, Martius O, Wanner H, Grebner D, et al. Weather patterns and hydro-climatological precursors of extreme floods in Switzerland since 1868. *metz*. 2012;21(6):531–50. <https://doi.org/10.1127/0941-2948/2012/368>
30. Stucki P, Bandhauer M, Heikkilä U, Rössler O, Zappa M, Pfister L, et al. Reconstruction and simulation of an extreme flood event in the Lago Maggiore catchment in 1868. *Nat Hazards Earth Syst Sci*. 2018;18(10):2717–39. <https://doi.org/10.5194/nhess-18-2717-2018>
31. van der Schrier G, Allan RP, Ossó A, Sousa PM, Van de Vyver H, Van Schaeybroeck B, et al. The 1921 European drought: impacts, reconstruction and drivers. *Clim Past*. 2021;17(5):2201–21. <https://doi.org/10.5194/cp-17-2201-2021>
32. Murphy C, Wilby RL, Matthews T, Horvath C, Crampsie A, Ludlow F, et al. The forgotten drought of 1765–1768: reconstructing and re-evaluating historical droughts in the British and Irish Isles. *Intl Journal of Climatology*. 2020;40(12):5329–51. <https://doi.org/10.1002/joc.6521>
33. Hawkins E, Alexander LV, Allan RJ. Millions of digitized historical sea-level pressure observations rediscovered. *Geoscience Data Journal*. 2022;10(3):385–95. <https://doi.org/10.1002/gdj3.163>
34. Craig PM, Hawkins E. Digitizing observations from the 1861–1875 Met Office Daily Weather Reports using citizen scientist volunteers. *Geosci Data J*. 2024;11(4):608–22.
35. Maugeri M, Bacci P, Barbiero R, Bellume M. Reconstruction of heavy rainfall events on the Southern part of the Alpine region from 1868 to the end of the 19th century. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*. 1999;24(6):637–42. [https://doi.org/10.1016/S1464-1909\(99\)00058-1](https://doi.org/10.1016/S1464-1909(99)00058-1)
36. Ufficio Centrale di Meteorologia Italiana. *Annali - Serie II. - Vol. IV. - Parte I*. Roma: Tipografia Sinimberghi; 1882.
37. Belluco E, Da Peppo L. La rotta dell’Adige a Legnago (Verona) del 1882 - alluvione del polesine e chiusura della rotta. *L’acqua*. 2022;5.
38. Biasillo R, Armiero M. The transformative potential of a disaster: a contextual analysis of the 1882 flood in Verona, Italy. *J Hist Geogr*. 2019;66:69–80.
39. Brönnimann S, Brugnara Y, Allan RJ, Brunet M, Compo GP, Crouthamel RI, et al. A roadmap to climate data rescue services. *Geoscience Data Journal*. 2018;5(1):28–39. <https://doi.org/10.1002/gdj3.56>
40. Allan R, Brohan P, Compo GP, Stone R, Luterbacher J, Brönnimann S. The international atmospheric circulation reconstructions over the earth (ACRE) initiative. *Bull Am Meteorol Soc*. 2011;92(11):1421–5.
41. Allan R, Endfield G, Damodaran V, Adamson G, Hannaford M, Carroll F. *Toward integrated historical climate research: the example of Atmospheric Circulation Reconstructions over the Earth*. Wiley Interdiscip Rev Clim Change. 2016;7(2):164–74.
42. Brohan P. Testing Google Vision for weather data rescue. Available from: <https://brohan.org/Google-Vision/>. Accessed 2025 July 22.
43. Brohan P. Testing AWS Textract for weather data rescue. Available from: <https://brohan.org/AWS-Texttract/>. Accessed 2025 July 22.