Data rescue in selected countries in connection with the EUMETNET DARE activity

Barbara Chimani1 | Ingeborg Auer1 | Marc Prohom2 | Mateja Nadbath3 | Anita Paul1 | Dubravka Rasol4

1Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria
2Meteorological Service of Catalonia, Barcelona, Spain
3Slovenian Environment Agency, Ljubljana, Slovenia
4Meteorological and Hydrological Service of Croatia, Zagreb, Croatia

Correspondence
Barbara Chimani, Zentralanstalt für Meteorologie und Geodynamik, Hohe Warte 38, 1190 Vienna, Austria.
Email: barbara.chimani@zamg.ac.at

Funding information
EUMETNET

Abstract
In situ data are an essential need in the analyses of past and current climate change. All over the world, different data rescue activities try to extend existing time series and to retrieve data in spatial and temporal data sparse regions. Also in Europe, part of the historic data is still only available in paper archives, without any digital access to these sources. The EUMETNET Data Rescue activity was started in order to investigate the potential for additional long-term stations, to monitor the progress of data rescue in European countries and to support participants with know-how exchange and international connections to facilitate data exchange. The activity raised the awareness of the need for data rescue and ongoing activities. This paper will provide information on the data rescue activities of 4 EUMETNET countries focusing on different aspects and in different states of the data rescue process. While Austria, Croatia and Slovenia are connected by a partly common history which is also displayed in the history of their meteorological network, Catalonia has a different historical background. Nevertheless, the problems data rescue is confronted with are similar for all activities. Therefore, the selected countries will provide a good overview of the problems and solutions discussed during the EUMETNET meetings. This paper wants to offer insight into different strategies for data rescue as well as into ongoing activities. Moreover, the importance of cooperation in regard to data rescue is underlined and examples of possible partnerships are provided.

KEYWORDS
Austria, Catalonia, Croatia, data rescue, digitization of historical data, EUMETNET, meteorological measurements, Slovenia

Information to data policy/access – restrictions: As the paper focuses on data rescue activities and not especially on datasets gained, no DOI number, specific information or link to the data are given. Slovenian data from 1948 onward are freely available at http://meteo.arso.gov.si/met/sl/archive/, while earlier data are available on request. Catalan data are freely available on request. Croatian and Austrian datasets are under regulation of the national meteorological services.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Geoscience Data Journal published by Royal Meteorological Society and John Wiley & Sons Ltd.
INTRODUCTION

Increasing interest in climate change and interpreting current changes in the light of past climate developments lead to an increased urgency of extending existing time series back into the past as far as possible (Kwok, 2017). This activity is also useful to improve existing reanalysis products and assess performance of climate modelling (Brunet & Jones, 2011). Therefore, data rescue (DARE) became an increasing priority in climate departments of National Meteorological Services (NMS).

DARE activities are now ongoing worldwide (e.g. northern Africa (Brunet et al., 2013), Asia (Williamson et al., 2015) or Australia (Ashcroft et al., 2016)), but it turned out that also in Europe there are still many data void spots when looking for long-term series. Several recent projects and initiatives all over Europe have driven DARE activities: EMULATE (European and North Atlantic daily to Multidecadal climate variability; Brunet & Jones, 2011, https://crudata.uea.ac.uk/projects/emulate/), ALOCLIM (Austrian Long Term Climate; Auer et al., 2001), HISTALP (Historical Instrumental Climatological Surface Time Series of the Greater Alpine Region; Auer et al., 2007, http://www.zamg.ac.at/histalp/) or UERRA (Uncertainties in Ensembles of Regional Reanalysis; Thorne et al., 2017, http://www.uerra.eu/) are some examples.

The topic is also in the focus of the World Meteorological Organization (WMO) and the Copernicus framework (European Union’s Earth Observations Programme), so that common activities between these two organizations evolved combining the experience of numerous parties experienced in data rescue in order to support less experiences parties. Examples of those are WMO/MEDARE (Brunet et al., 2013, 2014; www.omm.urv.cat/MEDARE/index.html) and ACRE (Compo et al., 2011; www.met-acre.org).

The main problems in common for most NMSs are few manpower, no external funding, insufficient information on their archives and missing equipment for digitization. Changes due to political reasons over the period of climate measurements at a number of stations cause additional obstacles in rescuing long-term time series.

The EUMETNET-DARE Activity (within the EUMETNET Climate Programme) started in 2013 with the goal to get an overview of possible improvements in the availability of long-term series and increase exchange between the European institutions in regard to data rescue. It brought together groups with different experiences focused on data rescue, including an exchange of methods as well as information on digitized data using communication platforms such as EUMETNET Data Management Workshops and the respective web portal (www.zamg.ac.at/dare). Particular attention was paid to identifying centennial series and those located in mountainous areas. A further development extended the exchanged information on foreign data available in the archives, in order to facilitate the task of investigating the history of old weather stations.

After 6 years of EUMETNET activity in this area, the focus on climate activities in this framework changed, but as the topic is still of high importance the exchange as well as climate data rescue is a continuing activity. Current information on ongoing data rescue activities is collected by the I-DARE activity (www.idare-portal.org; Brönnimann et al., 2019; Siegmund, 2014).

This paper gives an overview of the DARE activities performed in 4 of the countries during the EUMETNET activity, showing a band width of experiences. While the problems have been similar, the focus has been different in the four countries: Catalonia focused on the search for data from sources outside the archives of national meteorological services and the development of a tool for managing rescued digital information. Austria concentrated on the digitization and data quality control of historical climate data sheets stored in its archive. Slovenia focused on the search of data scattered in other meteorological services archives due to historical reasons and took up strategies from Catalonia. Croatia started its data rescue activities more recently and was able to profit from the experiences of the other countries.

HISTORICAL BACKGROUND

The development of station networks is mainly influenced by available financial resources and historical development. Figures 1 and 2 provide information on the evolution of the stations number in the four areas, and the dense network of observatories at the end of the 19th century. The political events also influence the management of climatological stations and, therefore, influences in which archives the data ends up. Figure 3 shows the great variability between the four countries, in terms of the institutions in charge of the observation networks. Table 1 gives an insight into the history of data rescue activities in the 4 countries.

DATA SOURCES

The first source for data are usually the archives of the national meteorological services, where, depending on the status of data rescue, they are available in printed or digital form. Data lost or unavailable in one country might be available in neighbouring countries or other archives due to data exchange or historical reasons.

Austria focused on the period of meteorological data being collected by national meteorological services (starting in 1851, with the foundation of ZAMG), and due to history, the national meteorological archive at ZAMG is the main data
FIGURE 1  Number of observing sites of existing meteorological networks ARSO (blue), DHMZ (green), SMC/AEMET (red) and ZAMG (black) in the course of time. The continuous line shows the number of manned stations, the dotted line the automatic sites. Between 1940 and 1946 there is a general decrease of stations because of the impacts of WWII, 4 years earlier in the case of Catalonia, as a consequence of the Spanish Civil War (1936-1939). High station density (peaks) about 1895 and 1970. Begin of automation between 1980 and 1990.

FIGURE 2  Station network in 1895/1896. The figure includes sites that reported mainly monthly - at some key stations daily or sub daily - reports. Sources are the Austrian Yearbooks (black), Hungarian Yearbooks (orange), Bosnian Yearbooks (green) and Yearbooks of the hydrological institute for Slovenian stations (blue). The Catalanian station network (red stations) contains at this time 34 measuring sites, mainly at the coast and some at islands like Mallorca and Menorca.
| Austria | Slovenia | Croatia | Catalonia |
|--------|---------|---------|-----------|
| upto 1851 | not organised by a single institution | | Granja Experimental de Barcelona |
| 1851 | | | Patxot’s pluviometric network, Sociedad Astronómica de Barcelona and AEMET |
| 1895 | | | |
| 1910 | | | |
| 1914 | | | |
| 1918 | | | |
| 1921 | | | |
| 1925 | | | |
| 1929 | Institute for Meteorology and Geodynamics in Ljubljana and under the Ministry of Construction & Hydrographic service of the Ministry of Public Works in Venice | | SMC and AEMET |
| 1939 | | | |
| 1941 | | | |
| 1945 | | | |
| 1947 | Hydro-meteorological Institute of Slovenia (part of the Federal Hydro-meteorological Institute of Yugoslavia) | | AEMET |
| 1991 | | | |
| 1996 | | | |
| 2001 | ARSO | DHMZ | SMC and AEMET |

**FIGURE 3** The figure shows the institutions managing the meteorological station network during history in Austria, Slovenia, Croatia and Catalonia (see abbreviation list at the end of this paper, for better understanding)
source. In contrast, Slovenia is one example of a country with historical developments that led to a great deal of climate records being stored in foreign archives, that is in Austria, Italy and Serbia. Therefore, within EUMETNET-DARE, the recovering of those climate records was the main focus of their data rescue activity. Along with the climate records, the recovering and rescuing of metadata was also part of their efforts. Inventories of meteorological stations with long-term datasets (48 stations), with possible long-term datasets (123 stations) and with meteorological stations, of which information might be available outside the country (approximately 150 stations), have been created. The list of Slovenian missing stations for the years before 1945 was created using printed, historical yearbooks as they offer information on climate data and metadata. In cooperation with other countries ARSO detected data from yearbooks between 1919 and 1946 stored in ISPRA (http://www.acq.isprambiente.it/annalipdf/), further data on Swiss webpage ‘Before 1921- Climate data rescue for the Southern Alps’ (https://before1921.wordpress.com/) and meteorological reports stored in the ZAMG archive. The meteorological reports of Slovenian stations found at ISPRA and ZAMGs archives were imaged for usage at ARSO. Additionally, copies of already known Slovenian data were found in historical yearbooks stored in Italy and Serbia. All mentioned lists on available and searched data as well as data from foreign countries in the different archives have been published on the web page of EUMETNET DARE (https://www.zamg.ac.at/dare/activities/lost-found). Croatia, starting their data rescue activities, was able to profit from the information from Slovenian colleagues about inventories and possible data sources from other countries. Therefore, part of their missing data could be located with little effort. These examples display the profit from and the growing need of inventories, sharing what each NMS conserves in its archives and making it available to other NMSs and research communities without hindrance.

Nevertheless, there are also other sources like newspapers that started publishing meteorological information in the early nineteenth century. The format and data available change over the time and differ between newspapers. The meteorological information published in the newspapers has advantages and disadvantages (Table 2), but it is a valid and helpful proxy

**Table 1** Information on data rescue activity

| NMS    | Activity                                                                                           |
|--------|---------------------------------------------------------------------------------------------------|
| ZAMG   | In 2008, ZAMG started the process of improving the archive and digitizing all available data. This activity is described in this paper. Before that, data rescue was only done in a limited way for special projects like ALCLIMP (Auer et al., 2001) HISTALP (Auer et al., 2007) |
| SMC    | SMC started data rescue activities described in this paper in 2005, when the historical meteorological archive was completely catalogued. Similar initiatives to the ones descript in the paper were already promoted decades ago by researchers at the University of Barcelona, with the reconstruction of the Barcelona climate series (Martín-Vide and Barriendos, 1995; Rodríguez et al., 2001) |
| DHMZ   | DHMZ started data rescue activities in 2018 but was involved in project like HISTALP (Auer et al., 2007) before that |
| ARSO   | ARSO started data rescue activities in 2005 including participation in international projects like FORALPS (http://www.ing.unitn.it/~foralps/, Dolinar et al., 2008) and MEDARE (http://www.omm.urv.cat/MEDARE/, Brunet et al., 2013, 2014). The data rescue activity described in this paper focuses on long time series |

**Table 2** List of Advantages (+) and disadvantages (−) of newspapers as data source

| Advantage | Description |
|-----------|-------------|
| +         | Many publications are currently digitized by historical archives and freely accessible via internet. As they are digitized following the Optical Character Recognition (OCR) method, this allows the rapid electronic search of keywords and the identification of potential newspapers containing weather data. Useful keywords can be the name (complete or abbreviated) of the most common meteorological variables or instruments (barometer, thermometer ...), always keeping in mind the language in which the publication is written |
| +         | Often the only reference to the existence of an ancient observatory, as the original observations may have been lost or destroyed |
| −         | Newspapers are an indirect source and may contain transcription errors (repetitions of values on consecutive dates are common) |
| −         | Temporary irregularity in the presence of the records often linked to the periodicity of the newspaper itself |
source in case of non-existence of original records. To find meteorological data within the newspapers, keyword identification is essential. Main information about metadata usually accompanies the data, that is location and observer, while other more detailed information is scarce (instrumentation or units). Nevertheless, there is hope to find metadata for data from newspaper as well, as the example of Slovenia shows: Most of the metadata to measurements for Ljubljana 1818-1856 published in newspapers had been rescued via documents imaged and published on the internet (Fromme, 1877; Lippich, 1834; Prosen, 2015; Reichsgesetzblatt, 1872).

In Catalonia, SMC focusing on data from the beginning of organized measurements in the region created a database (PREHISMET) of historical newspapers containing climate data. Public local, county and provincial archives with open access to their digital repositories are the main sources of information. By now, 202 publications have been identified, with data encompassing the period 1780-2010. Following the successful example of Catalonia, Slovenia took up the idea for data rescue from newspapers. So, old meteorological records for Ljubljana for the period 1818–1856 have been discovered in the newspaper ‘Laibacher Zeitung’ and digitized. On the internet, ARSO found 41 publications with historical climate records and/or metadata on Slovenian meteorological stations.

Moreover, in Catalonia, several campaigns have been undertaken to detect archives or repositories where original documents are likely to be kept, as historical data partly were not available in the meteorological service (Martin-Vide and Barriendos, 1995; Rodríguez et al., 2001). Private archives, mainly religious (Jesuits and Piarists) and corporative (e.g. hydroelectric power companies), have been catalogued in recent years, jointly with current notebooks of meteorological observers. An example of this activity is the agreement with ENDESA, the main Spanish electric company that transferred its meteorological archive to SMC: more than 22,000 sheets covering the period 1922-1992, containing meteorological and hydrological observations from sites located in the Pyrenees.

The amount of data rescued in the four countries during the EUMETNET activity from the different data sources is provided in Table 3.

### 4 DATA FORMATS

While today most of the data are collected automatically, directly processed by data quality control procedures and stored in databases, in the beginning of the meteorological services the observations were noted in climate diaries and copied on the pre-printed form of a climate sheet. Once established, the NMS also collected previous observational initiatives in their archives (i.e. made by naturalists, doctors or farmers, religious orders, universities or astronomic observatories). While for privately taken meteorological notes and publications in newspapers no fixed format was established, a standardization took place by NMSs (Figure 4). In the early years, the monthly climate sheet consisted of one page including basic metadata, the observations and measurements of the meteorological parameters air temperature and precipitation, air pressure, relative humidity, cloudiness, meteors, sunshine duration, wind speed and direction. Later, the climate sheet contained two pages. This format was adopted in many countries across Europe. Nevertheless, the format of climate sheets varied with time and region. The data were often published in printed year books, offering summarized information for all stations.

### 5 DIGITIZATION – METHODS AT THE NMHS

Depending on resources, digitization (both scanning and imaging) is mainly done by commissioning an experienced company or by the archives themselves. It is recommended to NMSs involved in DARE activities, to acquire at least a basic infrastructure for imaging (camera, tripod and lighting

| Number of stations total/ before 1948* | Austria ZAMG | Catalonia SMC | Croatia DHMZ | Slovenia ARSO |
|--------------------------------------|--------------|---------------|--------------|--------------|
| From national archives               | 799          | 833           | 823          | 131          |
| From newspapers                      | -            | 25            | -            | 1            |
| From other sources (religious, electric companies, military) | 44           | 31            | -            | -            |
| Additional sources (universities, academies) | 4            | 3             | 1            | -            |
| Number of stations (already digitalized) total/before 1948* | 433/195      | 820/302       | 226/32       | 410/190      |

**Table 3** Number of stations that were rescued in the framework of DARE, * before 1948 and 1941 for Catalonia resp., was defined as historical data due to political reasons (WWII)
equipment), which gives them autonomy, for example digitization of punctual documentary sources, or sources located in private repositories. In these cases, one or two people, with a basic knowledge about quality standards in imaging, can digitize a few hundred images daily (depending on the format), in an agile and cheap way (Figure 5). Table 4 provides information on the imaging procedure used. Table 5 on the priorities in the data rescue process, and Table 6 on the current status of data rescue.

At SMC, using the basic infrastructure described above, several internal projects of imaging meteorological documentation were carried out and many data from observers who keep weather notebooks or strip charts at home has being recovered. At ZAMG, an approach similar to that of SMC was used for imaging the Austrian climate sheets (Figure 5). The high resolution picture in jpg format was transformed into pdf and stored on a file server. A backup on tapes was performed periodically. Imaging of the climate sheets was performed manually by trained and qualified experts via a self-written software tool for conversion into adequate format. The keying was performed via DCT (Data Correction Tool, Table 7). In Slovenia, the keying of the logbooks and reports (historic records as well as those records that are still collected on paper today) is performed

FIGURE 4 Examples of data sources: upper right: first page of the digitized newspaper “Diario de Tarragona” corresponding to February 21, 1909, stored at SMC containing sub daily air temperature, air pressure, wind direction and sky conditions upper left: one-page climate sheet of Cven, stored and imaged at ZAMG, keyed and quality controlled at ARSO. In 1896, the climate sheets were printed in German. Provided is metadata (station name, date of measurement, station height, height of thermo- and pluviometer, observer’s name and observing times). Reported parameters have been air temperature (three times per day), cloudiness (three times per day), wind direction and velocity (three times per day), precipitation and atmospheric phenomena. lower: two-page historical climate sheet from 1872 of the site Senj (formerly Zengg) in Croatia printed bilingual in Hungarian and German - digitized at DHMZ. Additional to the information provided by the other shown climate sheet it includes information on the instrumentation and the parameters sub daily air pressure, wet-bulb temperature, relative humidity, vapour pressure and cloud categories
by trained and qualified experts and students, led by the experts. All keyed data are put into the same self-written database. The majority of climate records before 1948 is still in the paper archive. Part of them is keyed, but only Slovenian climate reports or logbooks from ZAMGs and ISPRAs archives are imaged, as ARSO, in contrast to SMC and ZAMG started keying from original documents. DHMZ started some additional efforts in data rescue, as part of the preparations for the project ‘Modernization of the National Weather Observation Network in Croatia – METMONIC’, which started in October 2017. Within the framework of the METMONIC project, three professional scanners (Figure 5) were bought in 2019. DHMZ carried out several pilot projects on getting a machine readable format from imaged data logs allowing for the lack of available man power involving students and DHMZ stuff (Table 8).

**TABLE 4** Digitalization, Data Quality and corrections tools at the institutes

|                        | Austria-ZAMG                                | Catalonia-SMC                               | Croatia-DHMZ                                 | Slovenia-ARSO                      |
|------------------------|---------------------------------------------|---------------------------------------------|----------------------------------------------|-----------------------------------|
| Imaging tool (hardware)| Camera fixed in a distance of 1 m above the imaging area | Camera fixed on a tripod (for in situ campaigns) and professional scanning machine (for wider collections) | Professional scanning machines: A3 and A2 book scanners for large formats and A3 high speed throughput scanner | External contractor’s A3 Flatbed Accessory Scanner |
| Digitization tool (software) | Self-developed software for digitization and conversion into pdf format | External contractor’s software for digitization and conversion into pdf/jpeg from TIF files | External contractor’s software for digitization and conversion into pdf/A format | External contractor’s software for digitization and conversion into pdf/A format |
| Data Quality control (man power) | 3 PM – educated staff | 2 PM – educated staff and 1 undergraduate grant-holder | 2 PM – educated staff | 2 - 3 PM – educated staff |
| Data quality control (software tool) (see Table 4 for more details) | DCT self-developed software | self-developed software and extraqc (Rclimdex, Aguilar & Prohom, 2011) | self-developed software | self-developed software |

**FIGURE 5** Digitization at the NHMs, upper left: camera at SMC, upper right: camera at ZAMG, below: book scanners from Zeutschel at DHMZ
For imaging, some points have to be kept in mind as follows: The digital storage space as the amount of images taken by SMC by 2020 (410,000 images of meteorological notebooks, strip charts, stations sites photos, scientific publications…) sums up to more than 12 TB, and the resolution of the image. In case of ZAMG, 600 dpi in jpg format was used. The main problem for digitalization are barely readable records. This can be caused by handwriting itself, bleaching ink and paper decay or because prior data quality controls made the original values nearly unreadable (Figure 6).

### Table 5 Information on priority in data rescue

| Institution     | Priority Details                                                                 |
|-----------------|----------------------------------------------------------------------------------|
| Austria-ZAMG    | Based on length, quality and measured parameters, giving highest priority to stations providing data for at least 30 years of all the main meteorological parameters in good quality (including well-documented meta-information), and high temporal resolution (daily). Additionally, stations with measurements before 1936 or during wartime were prioritized. Other stations have only been processed if they seemed important for spatial comparison to high priority stations |
| Catalonia-SMC   | Highest priority is given to long-term data and to stations located in mountainous and sparse areas. In addition, special attention is paid to series that can be linked to currently operating observatories, in order to create long-term series after applying a homogenization procedure |
| Croatia-DHMZ    | Highest priority is given to the long-term data especially the data from locations where there are still measurements. Other priority are meteorological phenomena lists from the stations for which the other data are already in the digital form. Third priority have precipitation data from regions where Standard Precipitation Index (SPI) is a very important information |
| Slovenia-ARSO   | Highest priority of digitization during the EUMETNET-DARE project was given to long-term data sets and its metadata. An exception are shorter datasets in mountainous and data sparse areas (50 years) or data needed for comparison or analyses of certain extreme weather event |

### Table 6 Status of data quality control at the institutions. DMHZ and ARSO did digitalization from original climate sheets directly and started with imaging of the sheets later on

| Institution     | Number of available climate sheets (total) | Imaged – from total | Keyed – from total | Quality controlled (QC) – from total | Quality controlled parameters |
|-----------------|-------------------------------------------|---------------------|--------------------|-------------------------------------|------------------------------|
| Austria-ZAMG    | 242,250                                   | 80%                 | 69%                | 52%                                | Everything on the climate sheet |
| Catalonia-SMC   | 661,000                                   | 87%                 | 65%                | 60%                                | Temperature (max, min, mean), precipitation |
| Croatia-DHMZ    | 3,691,800                                 | 2%                  | 50%                | 45%                                | Everything on the climate sheet |
| Slovenia-ARSO   | 1,171,280                                 | 2%                  | 70%                | 50%                                | Everything on the climate sheet |

6 | METADATA AND QUALITY CONTROL

Metadata can be found with the data itself as on climate sheets but are also included in postal correspondence between observers and meteorological services, where information on breakdowns, replacement of instruments and changes of observers among other information can be extracted. The SMC has been able to image dozens of letters with very useful information for quality control and homogeneity analysis. Additional sources might be instrument purchase invoices, inventories, sketches or photographs of meteorological stations, historical publications, etc.

For reading climate logbooks and yearbooks, especially in those regions that belonged to different political entities throughout their history, the knowledge of different languages is necessary as well as knowledge of Kurrent, Fraktur and Cyrillic alphabet. The attribution of data to the same station is complicated due to historical names of villages and towns which are partly changing due to changes in the official languages and policies (e.g. Kobarid (Slovene) – Caporetto (Italian) – Karfreit (German)).

Before 1871, other measurement units (e.g. Zoll, Paris Lines, Punkt, Inches, Reaumur) have been in use. Standard
observational times have not always been at 7°°, 14°° and 21°° (CET): from 1935 onwards, the measured amount of precipitation at 7°° CET has been assigned to the day of measurement, until that year it was assigned to the day before the measurement. In Austria, for example also an additional change of observation time from 21°° to 19°° CET in 1971 effected break points in time series. Transformation to today’s units and shifting the precipitation according to today’s standards have to be done before the data can be included in the database and is passed on to quality control. Shifts in observational times are usually not corrected before being stored in the database. Some information has to be provided on when a change took place and which observation times were concerned in which way.

Data quality control is the basis for climate research studies and analysis, based on long-term data in high temporal resolution. The first QC in DARE is visual during the keying process if done manually by inspecting the numbers on the sheet using meteorological knowledge. After this, first corrections based on metadata and a detection of gross errors,
the second step of QC is performed with the operational full-
or semi-automatic software packages (Tables 4 and 7). The
current status of QC of rescued data is provided in Table 6.

7 | DATA STORAGE AND ARCHIVING

While the keyed data can be stored in the usual database eas-
ily, the imaged data files and paper sheets need special treat-
ment. The ZAMG paper archive contains historical, up to
200 years old, climate sheets that are valuable cultural assets.
In 2012, the entire archive was restructured and the storage
was improved. The original documents are protected against
dust, dirt, humidity and light incidence. Using a self-written
visualizing programme (Figure 7), the images of the sheets
can be queried at the workplace with no need to go to the
archive.

In contrast, SMC had to look for alternative options as
in-house archivation is not possible. Therefore, an agreement
with the School of Conservation and Restoration of Cultural
Heritage of Catalonia (SCRCHC) and the National Archive
of Catalonia (NAC) has been signed. SCRCHC cleans and
restores the documents, NAC preserves them, and SMC is
responsible for digitizing and cataloguing the meteorological
sheets. ARTYDOC (Figure 7) is a digital tool that permits
storage, cataloguing and easily searching images for a better
consultation (Prohom et al., 2018).

ARSO being in a similar situation as SMC is cooperating
with the Archives of the Republic of Slovenia. Due to the
good experiences of Catalonia, they will follow the same path,
planning a digital archive after the example of ARTYDOC.
Therefore, before transferring the documents to the Archives
of the Republic of Slovenia, imaging all records (including
reports, charts, sketches...) is planned. Those digital files will
be stored in the digital archive for easier digitizing and qual-
ity control as well as for easy access within ARSO and to
preserve it in case of loss of the paper versions.

At the moment, DHMZ stores data sheets at several loca-
tions across Croatia which complicates approaching the data.
Moreover, not all locations are well suited as storage of paper
archives. As in 2020, the DHMZ has lost its headquarters
building due to the earthquake damages, the plan is to build a new building with space for an archive in accordance with Croatian National Archives (CNA) regulations.

8 | CONCLUSIONS

Archives are the monument and treasure of every nation; this is true also for the climate archives. Not only the climate information but also many indirect historical pieces of information are worth of analysing. The project EUMETNET-DARE rose the awareness of the importance to rescue and recover climate data and metadata. Depending on the national status of data rescue, this activity helped as motivation and guidance to get a better insight into the content of the archives or to start and increase data rescue activities. This was done by regular exchange, for example on examples of recovering and rescuing climate records, exchange on experiences with different imaging/digitalization methods, ways of QC and archiving.

Additionally, the close interaction facilitated the recovering of historic climate records stored in other countries and therefore the formation of long time series. This was especially the case for Slovenia and Croatia getting into contact with data rescue groups of countries with common historical periods. But it proved true that data can not only be found in those countries, where the data are expected, but in others as well due to exchange of yearbooks between different meteorological services. In this context, the EUMETNET-DARE web page has proven to be very beneficial. The WMO Guidelines on Data Rescue (WMO, 2016), as well as the International Data Rescue Portal (I-DARE, https://www.idare-portal.org; Brönnimann et al., 2019; Siegmund, 2014) proved to be helpful tools, too.

During the process of data rescue, different countries experienced individual obstacles, but also problems in common. Those have been of technical as well as of practical nature. The practical problems included the need of knowledge of different languages and written alphabets, difficulties in correct attribution of data with changing station names, and the use of different measurement units. While these were taken into account already during the keying and import to the database, considering other problems like changes in the observational standards (stations setup, changes in observational

**FIGURE 7** Digital archive tools upper: SMC-ARTYDOC. A record card of one of the first known meteorological notebook in Catalonia is shown: Dr. Salvà observations in Barcelona, starting back in January 1780. lower: ZAMG – climate sheet visualizing program. Query and view of the climate data sheet of Kremsmünster (the eldest station in our network, begin of measurements in 1764), including statistical output as number of climate sheets per year
times) and homogeneity will have to be tackled in forthcoming years.

The inspiring exchange on possible data sources, and solutions for technical problems (as between SMC and ARSO) helped to facilitate data rescue not only for those countries starting with data rescue and being able to profit directly from those ideas and experiences, but also for those being already deeply involved in data rescue, as it increased the visibility of the necessity and worth of the initiative.

9 | OUTLOOK

The activities of recovering and rescuing climate records and metadata will continue throughout all meteorological services. Slovenia and Croatia are planning to do imaging and organizing imaged documents in digital archives, digitizing, quality controlling, homogenizing and analyzing of climate records. Keying as a ‘citizen science’ activity, which already proved to be successful (Hawkins et al., 2019) is planned by SMC: The collection of weather sheets edited during the first years of the institution (1921-1939) will be released for digitization as part of the celebrations of its centenary. Austria will set its focus on imaging of not-yet-handled foreign data and optical character recognition (OCR) of the Austrian climate sheets and yearbooks including quality control with the aim to automatically transfer values into a database and to provide OCR subject indexing (search for keywords) for easier access for researcher and the community.

The newly digitized documents will improve quality control of historical data and will help to enlighten the knowledge on climate (e.g. extreme events, climate change) in areas for which a lack of climate records was given in the past (e.g. coastal areas and north-eastern part of Slovenia and Croatia).

ACKNOWLEDGEMENT

The historical climate records are the result of the persistent work of the enthusiastic, scientifically interested and accurate observers that we would like to honour. Further, we would like to thank all our colleagues that are continuously searching and collecting new data material, performing data quality control of the imaged and keyed datasets as source for climate change studies and analyses. Moreover, we are indebted to EUMETNET, who supported this activity. We also wish to thank the reviewers for their supporting comments.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally shareable data necessary to reproduce the reported results.

ORCID

Barbara Chimani https://orcid.org/0000-0003-0298-4181

REFERENCES

Aguilar, E. & Prohom, M. (2011) RClimDex-extra QC (EXTRA QC Quality Control Software). User Manual, http://etccdi.pacificclimate.org/software.shtml (accessed 20 April 2021).

Ashcroft, L., Allan, R., Bridgman, H., Gergis, J., Padmjenzy, C. & Thornton, K. (2016) Current climate data rescue activities in Australia. Advances in Atmospheric Sciences, 33(12), 1323–1324. https://doi.org/10.1007/s00376-016-6189-5

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R. et al. (2007) HISTALP – Historical instrumental climatological surface time series of the greater alpine region. International Journal of Climatology, 27, 17–46. https://doi.org/10.1002/joc.1377

Auer, I., Böhm, R. & Schön, W. (2001) Austrian Long-Term Climate 1767-2000. Österreichische Beiträge zu Meteorologie und Geophysik, 25.

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R. et al. (2019) Unlocking Pre-1850 instrumental meteorological records: A Global inventory. BAMS, 100, ES389–ES413. https://doi.org/10.1175/BAMS-D-19-0040.1

Brunet, M., Gilabert, A. & Jones, P. (2014) A historical surface climate dataset from station observations in Mediterranean North Africa and Middle East areas. Geoscience Data Journal, 1, 121–128.

Brunet, M. & Jones, P. (2011) Data rescue initiatives: bringing historical climate data into the 21st century. Climate Research, 47, 29–40.

Brunet, M., Jones, P.D., Jourdain, S., Efthymiadis, D., Kerrouche, M. & Boroneant, C. (2013) Data sources for rescuing the rich heritage of Mediterranean historical surface climate data. Geoscience Data Journal. https://doi.org/10.1002/gdj3.4

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X. et al. (2011) The twentieth century reanalysis project. Quarterly Journal of the Royal Meteorological Society, 137, 1–28. https://doi.org/10.1002/qj.776

Dolinar, M., Nadbath, M., Vičar, Z., Vertačnik, G. & Pavčič, B. (2008) Podnebni podatki v Sloveniji skozi zgodovino. Ljubljana: ARSO. Fromme’s Österreichischer Feuerwehr-Kalender, IV. Jahrgang, Str. 33–36 (Münzen, Maß, Gewicht), P. 33, Österreichische Nationalbibliothek, ANNO Historische österreichische Zeitschriften und Zeitschriften: http://anno.onb.ac.at/cgi-content/annoplus?aid=fwk&datum=1877&page=163&size=45

Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M. et al. (2019) Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. Geoscience Data Journal, 6(2), 160–173. https://doi.org/10.1002/gdj3.79

Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A. et al. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. International Journal of Climatology, 22, 1441–1453. https://doi.org/10.1002/joc.773

Kwok, R. (2017) Historical data: Hidden in the past. Nature, 549, 419–421.

Lippich, F.W. (1834) Topographie der k.k. Provinzialhauptstadt Laibach, in Bezug auf Natur- und Heilkunde, Medicinalordnung und Biostatik. Laibach, https://books.google.si/books/about/Topographie_der_k_k_Provinzialhauptstadt.html?id=hFZhAAAACAAJ&redir_esc=y
Martin-Vide, J. & Barriendos, M. (1995) The use of rogation ceremony records in climatic reconstruction: a case study from Catalonia (Spain). *Climatic Change*, 30, 201–221.

Prohom, M.J., Herrero, M. & Aparicio, A. (2018) ARTYDOC, a new digital archive of meteorological documentation. EMS Annual meeting, vol-15-259. [http://presentations.copernicus.org/EMS2018-259_presentation.pdf](http://presentations.copernicus.org/EMS2018-259_presentation.pdf)

Prosen, M. (2015) Profesor Friedrich Anton Frank – znan neznaneč, povzeto s spleta. [http://www.rad.sik.si/wp-content/uploads/2016/01/Frank.pdf](http://www.rad.sik.si/wp-content/uploads/2016/01/Frank.pdf)

Rodríguez, R., Barriendos, M., Jones, P.D., Martin-Vide, J. & Peña, J.C. (2001) Long pressure series for Barcelona (Spain). Daily reconstruction and monthly homogenization. *International Journal of Climatology*, 21, 1693–1704.

Siegmund, P. (2014) White paper on the International Data Rescue Portal (I-DARE). Geneva, Switzerland: WMO, pp. 16.

Thorne, P.W., Allan, R.J., Ashcroft, L., Brohan, P., Dunn, R.J.H., Menne, M.J. et al. (2017) Towards an integrated set of surface meteorological observations for climate science and applications. *Bulletin of the American Meteorological Society*, 98, 2689–2702. [https://doi.org/10.1175/BAMS-D-16-0165.1](https://doi.org/10.1175/BAMS-D-16-0165.1)

Williamson, F., Allan, R., Switzer, A.D., Chan, J.C.L., Wasson, R., D’Arrigo, R. et al. (2015) New directions in hydro-climatic histories: observational data recovery, proxy records and the atmospheric circulation reconstructions over the earth (ACRE) initiative in Southeast Asia. *Geoscience Letters*, 2(1), 2– [https://doi.org/10.1186/s40562-015-0018-z](https://doi.org/10.1186/s40562-015-0018-z)

WMO (2016) *Guidelines on Best Practice for Climate Data Rescue*. WMO No. 1182. 2016 edition, [https://public.wmo.int/en/resources/library/guidelines-best-practices-climate-data-rescue](https://public.wmo.int/en/resources/library/guidelines-best-practices-climate-data-rescue)

How to cite this article: Chimani, B., Auer, I., Prohom, M., Nadbath, M., Paul, A. & Rasol, D. (2022) Data rescue in selected countries in connection with the EUMETNET DARE activity. *Geoscience Data Journal*, 9, 187–200. [https://doi.org/10.1002/gdj3.128](https://doi.org/10.1002/gdj3.128)