# Identifying Data Rescue gaps and issues

C3S Data Rescue Service Deliverable Number: DC3S311a_Lot 1.1.2.1

By Manola Brunet[1, 2], Alba Gilabert[1], Phil Jones[2], Peter Siegmund[3]

**Delivered 16 September 2020**

**Summary of Contents**

This report builds upon an examination of currently available climate data in digital format kept in several global databanks. Using this information, we explore those regions, periods and variables that are underrepresented in the Copernicus for Climate Change Service/Data Rescue Service Portal. We illustrate ways to prioritise the recovery of those variables, periods and regions that are vital to support more robust climate applications, products, and services.

**Identifying Data Rescue gaps and issues**

Manola Brunet[1, 2], Alba Gilabert[1], Phil Jones[2], Peter Siegmund[3]

*[1] Centre for Climate Change, University Rovira i Virgili, [2] Climatic Research Unit, University of East Anglia, [3] Royal Netherlands Meteorological Institute*

## Table of Contents

## 1. Introduction

Under the Copernicus Climate Change Service (C3S) Data Rescue Service, we describe new efforts to coordinate and harmonise the many climate Data Rescue (DARE) activities around the world that are currently ongoing. The principal aim of these activities is to make in situ historical observations more easily discoverable, analysable, and available to a large variety of end users for further processing and analysis. Among the set of services to enhance the recovery of climate data worldwide, a new Data Rescue Portal was launched to garner and share information on DARE projects and the to-be-rescued data, which has been built and synchronised with the World Meteorological Organization (WMO)/Global Framework for Climate Services (GFCS) International Data Rescue (I-DARE) Portal.

The C3S DARE Portal eases the recovery of weather and climate data worldwide and has developed and maintains a global inventory of known DARE activities, providing a registry and access to standardised and searchable metadata on these activities. The Portal includes an interactive Map showing currently ongoing DARE projects (for which information is available) and the locations of the land surface stations, upper-air and marine observations that are part of the C3S Data Rescue Service Inventory. The inventory provides access to detailed metadata from the observations taken at stations located on the land-surface, from vehicles launched from fixed locations and moving platforms and from marine platforms. In addition, it also gives access to tools for scanning, digitising, formatting and quality control of both metadata and data, along with relevant guidelines and resources to assist in the recovery of relevant historical and present, but not-yet-digitised climate data, and guided by a distributed approach (Brönnimann et al. 2018). In short, it delivers a set of services that should ease the recovery of climate data, as well as making the rescued data more easily accessible through an upload service to the Global Land and Marine Observations Database (GLAMOD), a collaboration between the C3S and the National Centers for Environmental Information (NCEI) for assimilation into all reanalyses.

The C3S Data Rescue Service has its target to combine observations from different sources and archives worldwide to develop one comprehensive and usable dataset that contains all the observations found, providing an inventory of what is currently available or known to exist. The principal objective is to help and sustain ongoing endeavours to digitise the known historical observations and records and to make them more easily discoverable and accessible. The Service includes the tools, resources and guidelines developed to support all kinds of DARE projects, either small or large. In a nutshell, the C3S Data Rescue Service provides all the infrastructure to DARE practitioners, extending from databases, to software user support, community building, best practice guidelines and capacity development.

Those DARE activities are time- and resource-consuming processes, since they extend from discovering where old observations are archived, to imaging original data sources and their duplication and preservation, digitising the data and metadata contained in those sources and formatting the records into a standard digital format, usable for further analysis. All these stages require quality control as a vital and inescapable step to ensure all the observations transcribed and ingested into the databank are real measurements. Despite many recent efforts, vast amounts of data have not yet been digitised, including many past historical observations, and also many current measurements taken by manually-read instrumentation.

These vital records need to be unlocked through well-planned DARE exercises and to be made available to end-users for improving their weather and climate applications, assessments, and services (Brunet and Jones, 2011) and additionally to support climate or weather reconstructions (e.g. through reanalyses) based on instrumental data.

The C3S Data Rescue Service provides access through the Portal to inventories and other on-line resources (e.g. the inventory of early instrumental measurements) to updated information about finished, ongoing and planned DARE activities worldwide. In this regard, Brönnimann et al. (2019) provides, in addition to the global inventory of pre-1850 instrumental meteorological records, a detailed overview about global and national repositories, historical inventories, and other publicly accessible collections containing terrestrial meteorological observations to support data compilation and DARE efforts.

Gathering this kind of information is an essential step before prioritising the data to rescue in any DARE exercise. This, along with tools to visualise spatial and temporal data gaps in currently available digital climate data has been carried out by one of the partners of the C3S Data Rescue Service in co-ordination with the NOAA/CIRES/DOE Twentieth Century Reanalysis. This report is the focus of one of the many deliverables produced by this Service; namely, this deliverable DC3S311a_Lot 1.1.2.1 aimed at identifying gaps in data availability to prioritise DARE projects.

Therefore, this report will provide insights into the spatial and temporal gaps in the availability of climate records in digital format from some of the larger and publicly available global data repositories. The emerging data gaps from these repositories will be checked against the holdings of the to-be-rescued data in the Portal inventories with respect to regions, periods and meteorological variables, together with the steps to help prioritisation in future data rescue activities. In addition to the data gaps assessment, an overview of problems and issues with Data Rescue will also be briefly addressed.

Besides this introduction, the report is organised as follows: the second section provides an overview of the availability of climate data in digital format from global datasets, while the third section discusses the contents of the C3S Data Rescue Portal. Finally, in section four, a potential prioritisation of DARE projects is provided.

## 2. Overview on climate data availability in digital format from global datasets

There are several publicly accessible global data repositories containing climate data in digital format, composed of either in-situ observations or derived products, such as gridded data and reanalyses products. Their development supports climate data analysis whilst also serving a wide range of other purposes and applications. Among others, there are the temperature datasets curated by the Climatic Research Unit (CRU) at the University of East Anglia (Norwich, the UK), such as the HadCRUT4, a global temperature dataset containing gridded monthly temperature anomalies all over the world taken both at the land and marine surface that have been used to document long-term temperature change at the global and hemispherical scales. With a similar purpose there is the Global Historical Climatology Network (GHCN) from the National Oceanic and Atmospheric Administration (NOAA)/National Center for Environmental Information (NCEI) containing an integrated database of air temperature measurements (monthly and daily scales) subjected to quality control and homogeneity assessment. With a focus on precipitation, there are also a number of global datacentres providing access to in-situ and gridded products at the monthly and daily scales to support climate monitoring. Among them are CRU, GHCN and also the Global Precipitation Climatology Centre (GPCC). This is run by the Deutscher Wetterdienst (DWD) under the auspices of the WMO to provide global precipitation assessments based on gridded gauge-analysis products. CRU and GHCN provide similar products and also make their station data available.

For a different meteorological variable, air pressure, the International Surface Pressure Databank (ISPD: Cram et al. 2015) provides the world's greatest assemblage of pressure measurements spanning 1722 to 2015 and incorporating both station-based and marine observations, along with tropical cyclone tracks as necessary input data to Reanalyses. These data recovery efforts have been facilitated by the Atmospheric Circulation Reconstructions over the Earth (ACRE: Allan et al. 2011) Initiative. For example, ISPD data feed into the Twentieth Century Reanalysis (20CR) Project (Compo et al. 2011, Slivinski et al. 2019) to enhance historical reanalyses that span more than a century and support a wide range of climatic analyses, such as climate trend assessments or diagnostics of the impacts of individual extreme weather events.

Despite the massive amount of climatic data available in digital format, there are large volumes of observations taken both in the past and at the present that still remain locked and are unusable to support any climate assessment or the delivery of any climate product and service Examining data gaps in time and space in the available climate data in digital format is not an easy task, due to the diverse number of data repositories and their various formats and purposes. Catalogues of the input data in these international databanks are not always available or easily examinable. In addition, to prioritise those records taken in areas and periods in more urgent need of being rescued, there is the need to explore first the availability of climate data in digital format from those global datasets currently available.

Therefore, to contribute to enhance the usability of the rich heritage of climatic data, the C3S Data Rescue Service is also channelling tools and resources to identify data gaps in time and space by using the 20CR fields, as already mentioned in the introduction to this report. In this regard and supported by the Oldweather initiative, Phillip Brohan, one of the partners of the C3S Data Rescue Service consortium, has developed several visualisation tools that enable the available digital data to be mapped using version 3 of the 20CR for the period 1850-2011.

In this regard, the coverage of weather observations (i.e. near surface air temperature, precipitation values, 10 m wind fields) and their uncertainty has been assessed from the 20CRv3 first ensemble member for 1x1 degree grid-cells in an animated video for the aforementioned period. Locations with observations in the 1x1 grid of the 20CRv3 and full details on data source and construction methods can be visualised at https://oldweather.github.io/20CRv3-diagnostics/obs_video/obs_video.html and https://vimeo.com/364280156 both showing the effects of limited data, while weather conditions across time can be seen here: https://oldweather.github.io/20CRv3-diagnostics/fog_videos/time_skip.html. The lack of observations on the simulated climate conditions throughout time is clearly visible in the animated map, where the globe is covered by fog (the fog of the ignorance) due to no information until around 1950 but improving afterwards, although with still many foggy regions particularly in Central Africa, parts of South America and most of the Southern Hemisphere oceans. These visualisation tools provide insights on data availability and, therefore, their gaps throughout the 1850-2011 period and allows easy identification of those regions and periods with the worst coverage.

In addition, the C3S Data Rescue Service has produced other resources that allow DARE practitioners to identify data gaps and know where there might be sources to fill them. This is the case for the compilation and assessment carried out by Brönnimann et al. (2019), aimed at providing a global inventory of early instrumental observations and their sources for the pre-1850 period. This pointed to the high potential for undertaking new DARE projects, but also to a few datasets that can no longer be found. The full recovery of this rich heritage of climate data would allow the extension back in time of the climatic record to support the development of derived data, such as reanalyses products, along with analysis of climate variability on centennial time scales. This community effort has unlocked thousands of non-digitized climate series, providing details about where, when and by whom meteorological observations were recorded prior to 1850 and where their sources are located (see Table 3 of Brönnimann et al. 2019 listing publicly accessible data repositories containing pre-1850 measurements).

Starting from the past to the present, the entries in the inventory from Brönnimann et al. (2019) indicate that nearly all early and long series were taken in Europe, although there are also early series in New England and Canada, while remaining parts of the world only show sporadic or short early records. The stocktaking includes 4,583 unique entries for about 2,250 locations. The first series begin circa 1650 and from 1680s onward there are at least 10 records inventoried each year, rising to about 50 in 1720 and reaching hundreds by 1800. By 1800 coverage had expanded, and although scarce, has reached all regions except Africa, which along with the Artic has reasonable coverage only by the period 1851-1890, while no entries were identified for Antarctica until the 1890s or early 1900s. In 1850 more than 1200 entries exist globally, with half of these taken over Europe. More than 400 are in North America, Central America, and the Caribbean, while over Africa, South America, and South-west Pacific regions less than ten entries were found pre-1850. It is important to note that only around 20% of early years (in Brönnimann et al., 2019) have been transcribed and are available from global repositories. This clearly indicates that a large fraction of the entries are not used into the global archives, highlighting the need for their transcription. As concluded by their authors, expanding data series back in time requires further efforts, which includes a more comprehensive metadata listing, along with full inventorying and updating. Often data might be available as monthly averages, but daily or sub-daily timescales need digitising. Last but not least, continuing imaging, preserving and transcribing more early instrumental data sources, along with compilations of their metadata for digitally available data in a common repository should be pursued.

From the data gaps visualisation exercise undertaken by Philip Brohan mentioned above for post-1850 data, several data sparse regions and periods emerge that should be focussed upon as targets for future DARE projects. Building mainly upon the land-surface station data and ship log-book observations, an animated temporal evolution of the coverage of these observations from 20CRv3 can be gained and visualised at https://oldweather.github.io/20CRv3-diagnostics/obs_video/obs_video.html. From exploring observational availability through time, the best level of observational coverage can be seen globally since the late 1950s onwards, either over land (stations data) or marine surface (ship log-books), although even in this generally good period, there are still regions and subperiods with data sparsity. The level of coverage pre-1950s is mainly characterised by the scarcity of observations both over the land and the sea, particularly from 1850 to 1920, when there are relatively few observations over Europe, parts of Asia and North America, and marine observations were mostly from the Atlantic ocean, especially the North Atlantic Basin.

It can be seen that there are scattered observations over Europe and North America, especially its Eastern part, and almost no information over the Pacific Ocean with few ship-tracks reporting data over the Atlantic and Indian Oceans for the period 1850-1880. From 1880 to 1920 increasing observations can be seen over Europe, Asia, and the contiguous United States (US), with scattered observations over Australia and a very small number of coastal locations in Africa, South America, and Eastern Asia.

Improved coverage over the marine surface is reached from the 1920s onwards, when also the Pacific and Indian Oceans show more ship-tracks, although the dominant coverage is mainly over the Northern Hemisphere (NH) marine surfaces. It will not be until the late 1930s when a denser coverage over land surfaces can be highlighted, particularly over central Europe and Eurasia, the contiguous US, while scattered observations are evident over the eastern part of Australia and the south of South America. However, there are large gaps over China, western and south-eastern parts of Asia, the Arabian Peninsula, India, southwest Europe, Middle East, Africa, Central America, and the north of South America. The poor coverage shows mainly over land and marine surfaces of the Southern Hemisphere (SH) and this is not improved until the mid to late 1950s, when observations from China, Southeast Asia, and parts of Africa are added. While the marine surface is better covered at that time, not only in the NH but also over SH seas, still many spatial gaps appear over land. This is particularly true for observations over southern and western Asia, including India, the Arabian Peninsula and central America for the 1960s. A dramatic increase in coverage from the mid-1970s onwards occurs over some of the previously data-sparse regions; namely, the Siberian region along with locally scattered observations over the Artic Basin and Canada, northern, central and southern parts of Africa, the whole Australia, New Zealand, South America, Mesoamerica and the Caribbean, and coastal observations around Antarctica. However, in the case of African land observations, the level of coverage varies erratically from the 1970s onwards, with intermittent improvements and deteriorations.

In summary, the information emerging from the visualisation tools developed by Phillip Brohan point to the need to improve the observational coverage over the SH and the tropics, both over land and marine surfaces, particularly for the second half of the 19th and early 20th century in the case of the marine observations and for the second part of the 19th century and the first half of the 20th century in the case of land stations, especially over Africa, southern and western Asia, south and central America and the northern and southern high latitudes.

Finally, the development of similar data visualisation tools and even the production of more friendly dataset catalogues should be pursued and applied to other climate datasets since the different global databases are not exactly similar and many differences among the data series they contain could be important in some regions.

## 3. The C3S Data Rescue Portal and its assets

The C3S Data Rescue Service Portal sustains a database synchronised with the I-DARE Portal that contains searchable information about all kind of DARE projects, either small or large. Information on DARE projects includes both details on the data to be rescued (e. g. domain, variables, their time-steps, period of observation, locational details) and on the specific project (e.g. organisation responsible, contact details, project status, links to the project). About 124 DARE projects are fully or partially documented in this database and Figure 1 gives their locations.

DARE projects are spread across the globe, although there are regions with higher number of projects than others. It has to be taken into account that the DARE projects are assigned to the WMO region from where the DARE activities are carried out, but this doesn't always mean these activities have the focus on the region they have been assigned to.

As shown in Figure 2, the region with the largest number of DARE projects is Europe with 30% of the total. Next is Africa with 21%, where the greatest efforts to support DARE projects have been made. Asia and South-West Pacific account for 16% and 15% respectively and the rest of the regions represent less than 10%; namely, North and Central America and the Caribbean with 8%, South America with 6%. Finally, there are some projects focused on retrieving data from different regions, which represent 3% and other projects that are global with 1%.

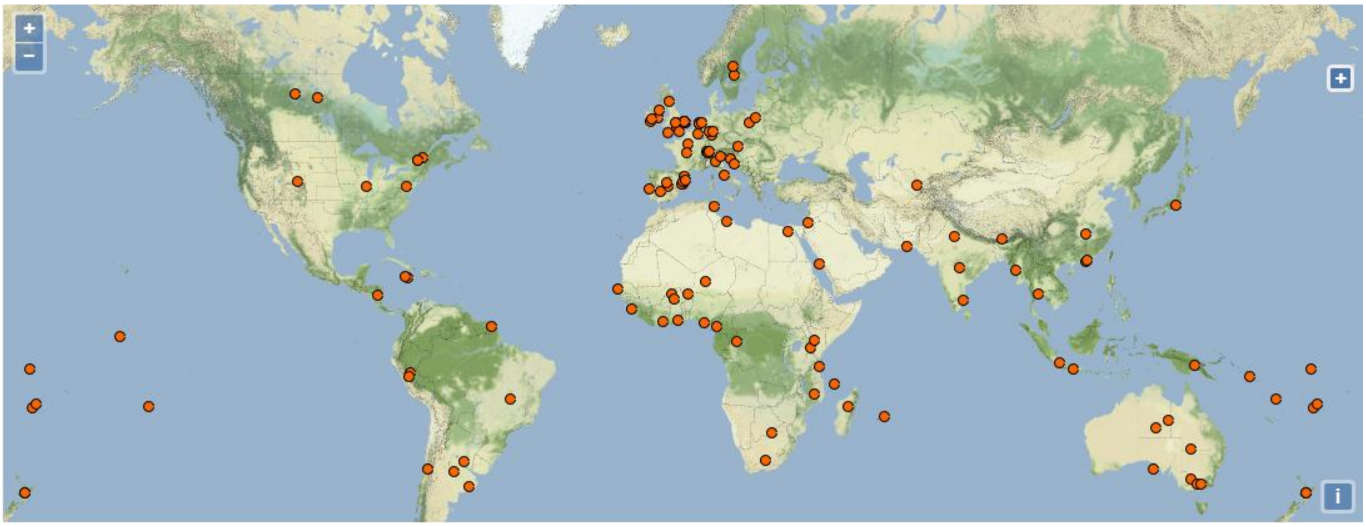

Figure 1. Image of the I-DARE interactive map showing locations (red dots) of currently available Data Rescue projects (124). Source: https://datarescue.climate.copernicus.eu/map

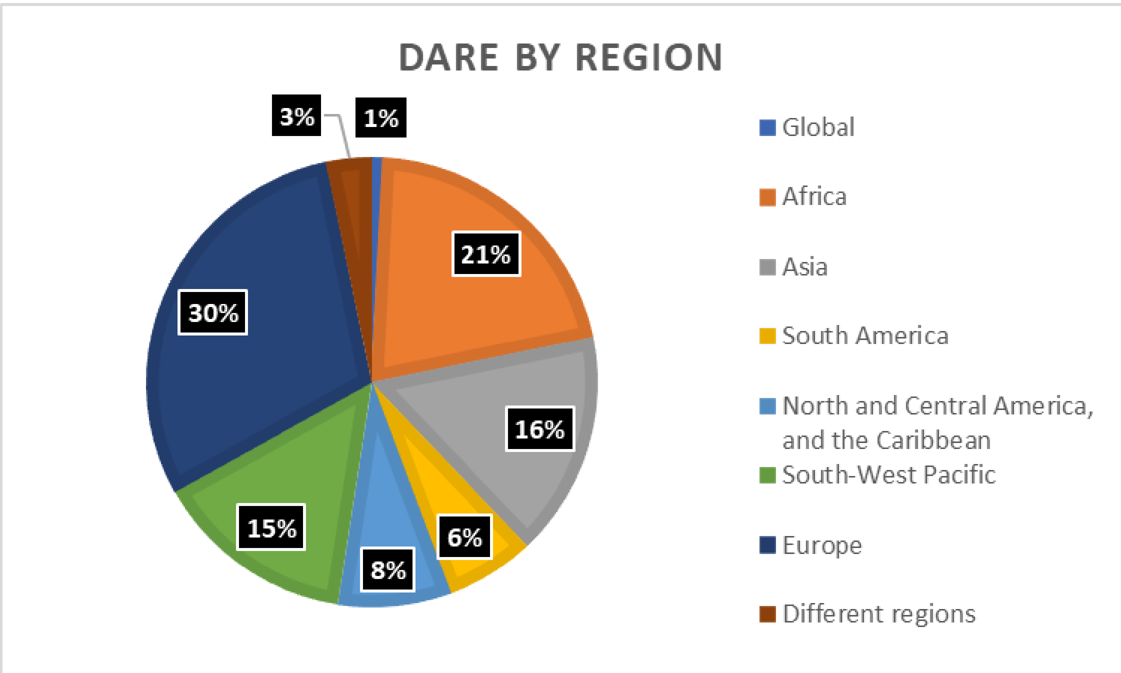


Figure 2. Pie chart showing the proportion of DARE projects in each of the World Meteorological Organization Regional Associations

Regarding the domain to which the data to be recovered belongs, Figure 3 shows their percentages, indicating that the atmospheric domain is the one that attracts more effort, followed by the marine domain. So, more effort has been paid to rescue meteorological data taken over land surface than marine, or most of the documented DARE projects in the Portal have an atmospheric focus. There are, though, a high fraction of DARE projects for which the targeted domain is not given.

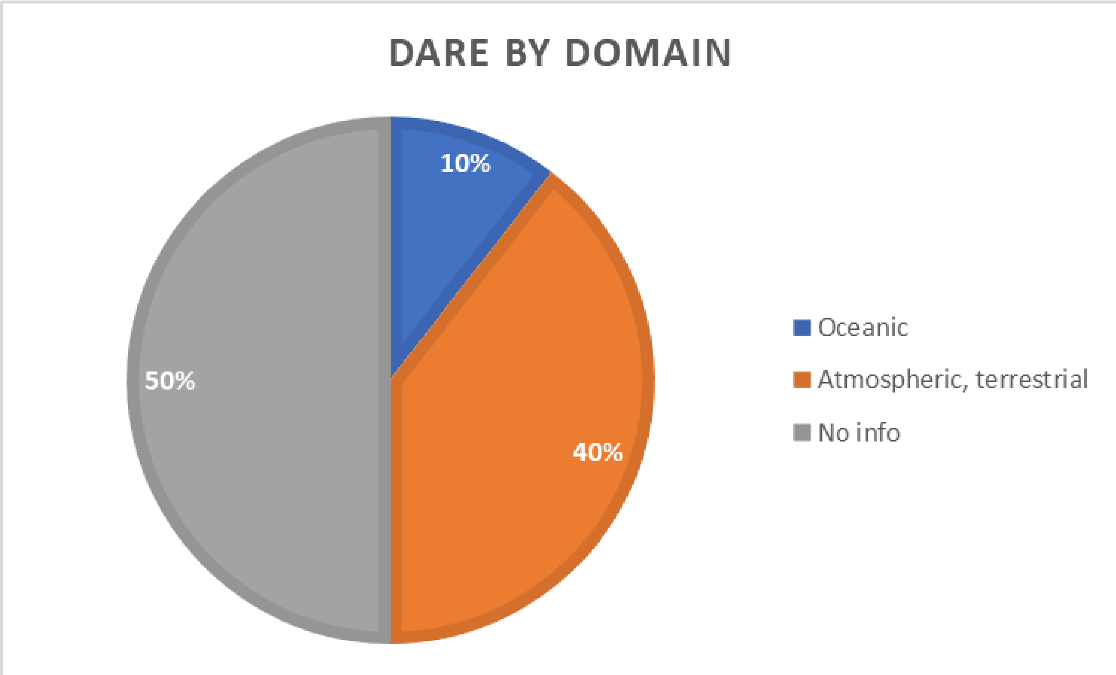


Figure 3: Pie chart showing the proportion of DARE projects by domain (oceanic, terrestrial or no information)

In the case of those climate variables that attract more interest to be rescued (Figure 4), temperature and precipitation data have been the most popular variables to be rescued, followed later by wind, air pressure, and humidity with much less interest. The most widespread use has been made of pressure, as the Reanalyses are widely used. However, there is a remarkable number of the documented projects in the Portal that do not provide information about the variables to be rescued. The recovery of data from other climatic variables different to temperature and precipitation are likely due to the needed as input to reanalyses projects.

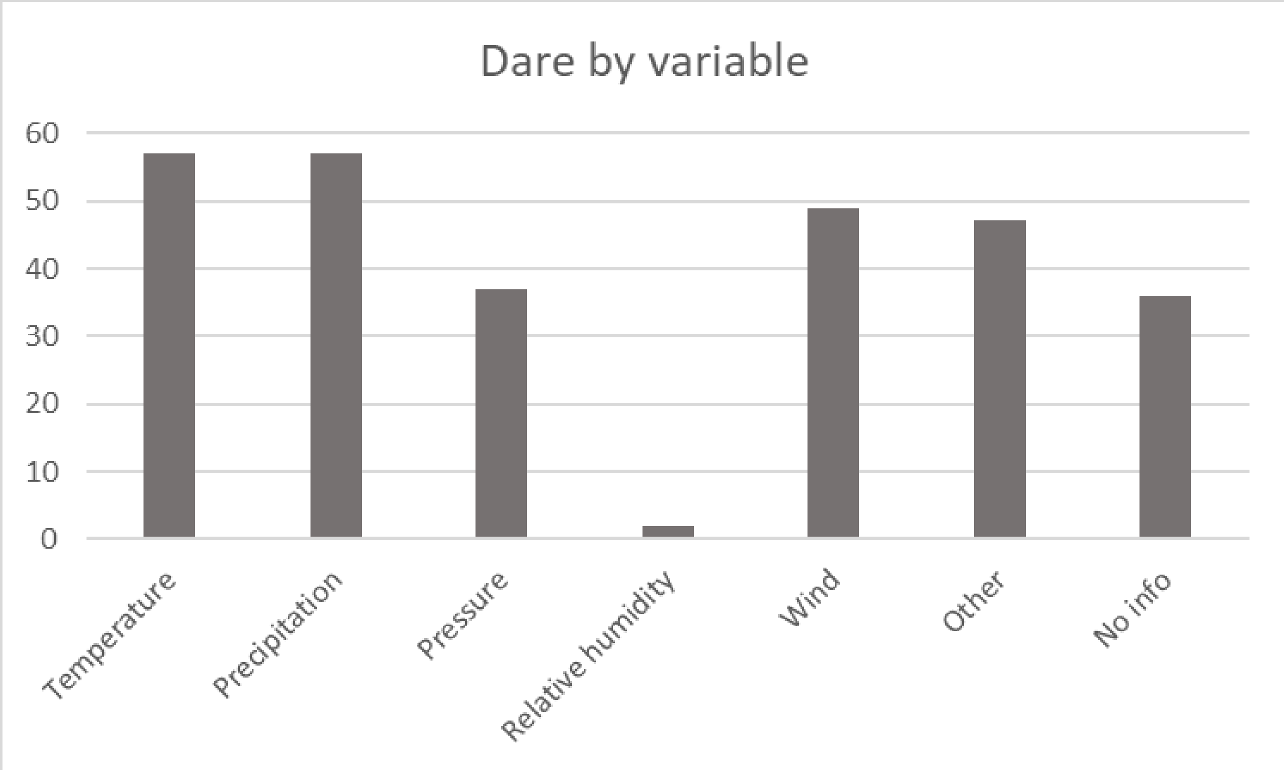


Figure 4. Number of the DARE projects documented in the C3S Data Rescue Service Portal by climatic variable target to-be-rescued

About the time periods that have attracted more DARE efforts, Figure 5 gives the number of DARE projects providing their availability (left plot) and the number of projects per century (right plot). The 20th century, followed by the 19th century have been the two centuries where more DARE projects have focused their recovery efforts. The 21st and the 18th centuries have also attracted interest, with the 17th century being the period with the least projects. When examining by regions those periods with the largest DARE efforts (not shown), this indicates the efforts paid to recover data over Africa and Europe for the 19th and 20th centuries, while over Southwest Pacific and North and Central America and the Caribbean (Asia and South America) it is the 19th (20th) century. It must be highlighted, though, the fraction of projects providing information about the period for which the data have to be rescued is very low (Figure 6), since only 30% of the documented DARE projects have complete information.

Current status of DARE projects is provided in Figure 7, but only 31% of the documented projects in the Portal give this information. Most of them are ongoing (19%), with lower percentages for those that have ended (5%), while the remaining fraction (7%) are onhold, postponed or planned.

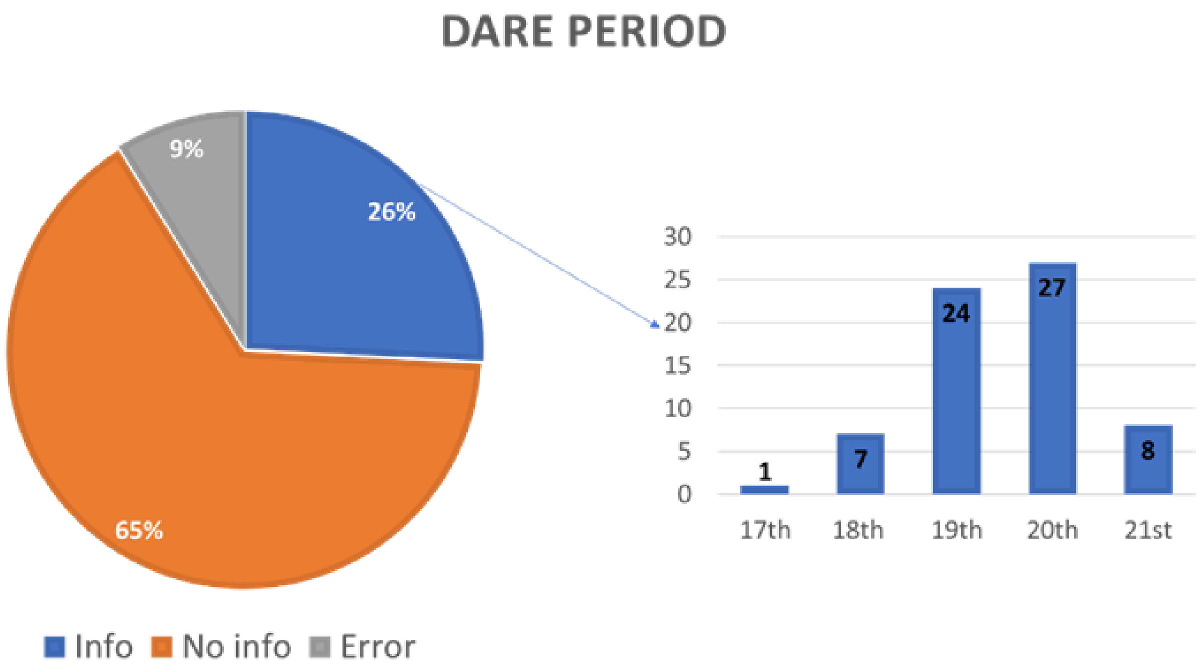


Figure 5. Number of DARE projects by century (right plot) and percentage of projects providing information about the period to which the data to be rescued belong to (left plot).

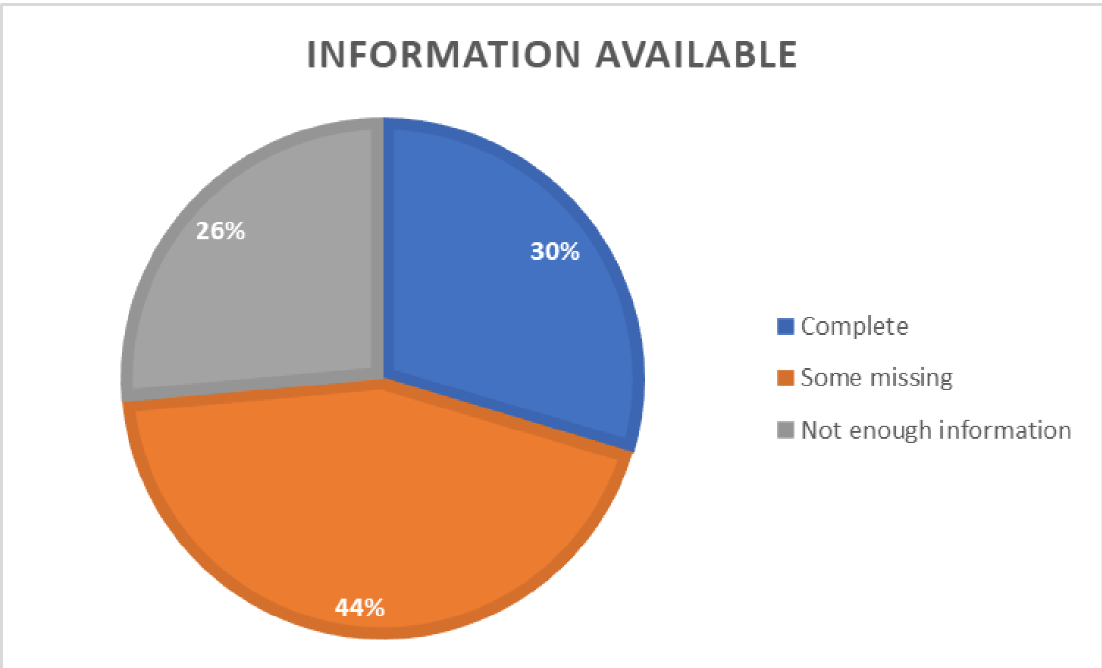


Figure 6: Pie chart showing the completeness of the projects available in the DARE projects Portal

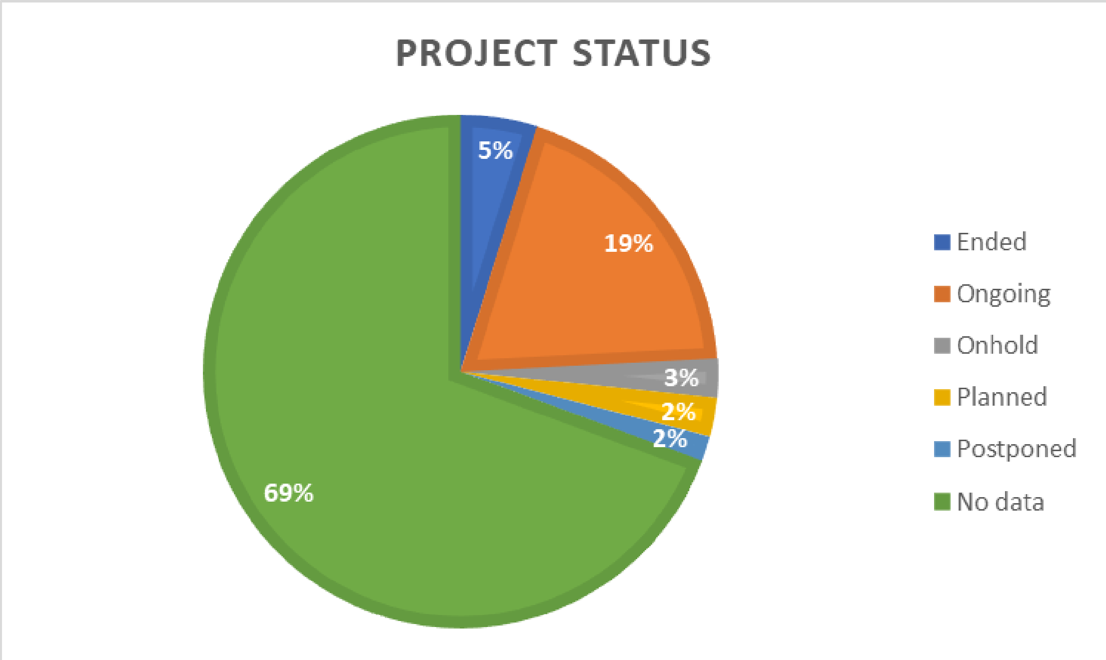


Figure 7. Pie chart providing the status of the DARE projects documented in the Portal

In summary, a remarkable number of projects have some details in the DARE Portal, but not all of them are fully documented, which limits its usefulness for prioritizing either regions or periods to be rescued from currently available information. Therefore, more effort has to be paid to properly document the DARE projects available through the Portal, as well as the data available in the metadata inventories of the C3S Data Rescue Service, since this currently contains only the information gathered by six projects. The currently documented, either partially or fully, DARE projects is a first step to help global coordination of DARE activities, but the contents of the Portal have to be improved. This is also crucial to the promotion of date rescue and to a large number of DARE's practitioners providing their input into the Portal, input that should be guided when being introduced to avoid current mistakes (e.g. reporting about the duration of the project, instead of the period for which the data are being recovered). Also the implementation of basic quality controls about the information provided by each project should be envisaged.

## 4. Prioritising Data Rescue targets and its issues

To prioritise the undertaking of DARE projects requires having an adequate assessment of what is available in digital format at least in global climate data banks and repositories. But this is not a trivial task, since for many of the currently accessible databanks a comprehensive catalogue of the data they contain is not often easy to find nor easy to analyse. The development of data catalogues should be an activity envisaged by all databank managers and, perhaps even more important, the search for resources to crosscheck and link the metadata of all global datasets in a unique repository, such as the C3S CDS, should be pursued and supported. Despite this difficulty, detailed and up-to-date information on digitally available data has been provided in section 2 by means of exploring the data available in the ISPD across time, which has been made possible thanks to the tool developed by Phillip Brohan in the context of the C3S Data Rescue Service contract.

In addition, before prioritising any DARE exercise, it is vital to know what DARE practitioners are undertaking at present or plan to in the near future. To compile this kind of information is the rationale behind the implementation of the C3S Data Rescue Service Portal, where metadata about DARE projects is documented and the metadata inventories for the different observing domains are kept. However, the information gathered is far from complete, and is difficult to use for prioritising DARE projects. This is likely due to the fact that DARE practitioners introducing the information about their projects in the Portal have not been properly filled the information fields to avoid mistakes being introduced, such as those mentioned in the third section. In addition, to upload information into the Portal databases is not a trivial task, particularly when the introduction of metadata about the data to be rescued requires much work that few DARE practitioners have time to do. Therefore, to derive resources to support not only its implementation but also its amelioration and continuation into the future, support by funding agencies is essential. Despite these drawbacks, a summary of the already available information on the to-be-rescued data in the portal has been provided, which along with experts and users of climate data allows us to give insights about the regions, periods and meteorological variables to rescue in future DARE exercises.

Nevertheless, to prioritise which data must be recovered first, it is also important to consider what is the purpose of the application or the analysis the data are needed for. There are so many different application purposes that makes the prioritisation task difficult, but at least DARE activities are now well-known. They are not as well supported as they should be though, as they rely on the goodwill of volunteers and this will probably never bring this activity to a successful end or to the full recovery of our rich heritage of meteorological data.

One of the end goals of many recent DARE projects has been to provide data to be assimilated into the global and regional reanalyses. The reanalyses input data has extended back in time as far as possible those global reanalyses products by providing them with observed data series that are missing. Reanalyses will always need to be periodically rerun, not just because the models improve and computers get faster, but because more DARE projects produce more vital input. It is important to combine these two drivers of reanalysis. At lower spatial scales (e.g. regional and local scales) the need lies to improve the input data over those data-sparse regions with few or not digitally data available during the last decades. Many other applications also need more and better input data to be carried out (e.g. climate monitoring products, climate variability analysis, seasonal forecasts, or climate prediction studies). Placing the focus on data supporting assessments of climate variations and trends, for instance, would require rescuing more measurements taken over Africa, parts of Central and South America and Southeast Asia at the monthly scale from the mid-19th century to the mid-20th century. Unlocking observations from ships logbooks that sailed in the southern oceans, especially the Pacific Ocean during the late 19th century and early 20th century, would be also required. Issues here would be whether there are any data to digitise in these areas or not and whether the potential data sources are imaged or not. Therefore, it should be envisaged to count on a complete list of relevant data sources, sorting out their images and building a catalogue of the data they contain as a prerequisite for any successful DARE effort.

We propose to focus on two priorities: (1) recovering air pressure data to feed global reanalysis products and (2) recovering temperature data to support assessments of climate variations and trends. To feed global reanalysis products require enhancing their input data, both those observations taken on land and marine surfaces, at the hourly scale over several periods and regions showing data scarcity, pointing to the need to enhance observational availability over the SH. For land surface data, the whole of Africa, particularly its eastern and central parts, part of Central and South America, along with Southeast Asia and the Western Pacific should be prioritised to recover pre-1950 observations, while for the marine surface, pre-1920 observations taken by ships in the Southern Oceans should be placed on the focus for their recovery, to extend back in time and make more robust the historical global reanalyses. The assessments of climate variations and trends require rescuing more measurements taken over Africa, parts of Central and South America and Southeast Asia at the monthly scale from the mid-19th century to the mid-20th century. Unlocking observations from ships logbooks that sailed in the southern oceans, especially the Pacific Ocean during the late 19th century and early 20th century, should also be an important focus.

**References**


Allan Rob; Philip Brohan; Gilbert P. Compo; Roger Stone; Juerg Luterbacher; Stefan Brönnimann (2011). The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative, Bull. Amer. Meteor. Soc. (2011) 92 (11): 1421–1425. https://doi.org/10.1175/2011BAMS3218.1

Brönnimann S., Yuri Brugnara, Rob J. Allan, Manola Brunet, Gilbert P. Compo, Richard I. Crouthamel, Philip D. Jones, Sylvie Jourdain, Jürg Luterbacher, Peter Siegmund, Maria Antónia Valente, Clive W. Wilkinson. 2018. A roadmap to climate data rescue services. Geosciences Data Journal, Volume 5, Issue1, June 2018, pp.28-39

Brunet, M. Jones, PD. 2011. Emerging data rescue initiatives: bringing historical climate data into the 21st century. Climate Research, DOI:10.3354/cr00960

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X.L. Wang, S.D. Woodruff, and S.J. Worley, 2011: The Twentieth Century Reanalysis Project. Quarterly J. Roy. Meteorol. Soc., 137, 1-28. DOI: 10.1002/qj.776.

Cram, T.A., G. P. Compo, X. Yin, R. J. Allan, C. McColl, R. S. Vose, J.S. Whitaker, N. Matsui, L. Ashcroft, R. Auchmann, P. Bessemoulin, T. Brandsma, P. Brohan, M. Brunet, J. Comeaux, R. Crouthamel, B. E. Gleason, Jr., P. Y. Groisman, H. Hersbach, P. D. Jones, T. Jonsson, S. Jourdain, G. Kelly, K. R. Knapp, A. Kruger, H. Kubota, G. Lentini, A. Lorrey,

N. Lott, S. J. Lubker, J. Luterbacher, G. J. Marshall, M. Maugeri, C. J. Mock, H. Y. Mok, O. Nordli, M. J. Rodwell, T. F. Ross, D. Schuster, L. Srnec, M. A. Valente, Z. Vizi, X. L. Wang, N. Westcott, J. S. Woollen, S. J. Worley, 2015: The International Surface Pressure Databank version 2. Geoscience Data Journal, 2, 31-46. DOI: 10.1002/gdj3.25.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez‐Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K. and Wyszyński, P. (2019), Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. Q J R Meteorol Soc. (accepted) doi:10.1002/qj.3598.

**Acronyms:**

| | |
|---|---|
| 20CR | Twentieth Century Reanalysis |
| ACRE | Atmospheric Circulation Reconstructions over the Earth |
| C3S | Copernicus Climate Change Service |
| CDS | Climate Data Store |
| CIRES | Cooperative Institute for Research in Environmental Sciences |
| CRU | Climatic Research Unit |
| DARE | Data Rescue |
| DOE | United States Department of Energy |
| DWD | Deutscher Wetterdienst |
| GFCS | Global Framework for Climate Services |
| GHCN | Global Historical Climatology Network |
| GLAMOD | Global Land and Marine Observations Database |
| GPCC | Global Precipitation Climatology Centre |
| I-DARE | International Data Rescue (Portal) |
| ISPD | International Surface Pressure Databank |
| NCEI | National Centers for Environmental Information |
| NH | Northern Hemisphere |
| NOAA | National Oceanic and Atmospheric Administration |
| SH | Southern Hemisphere |
| US | United States |
| WMO | World Meteorological Organization |