



PROJECT DELIVERABLE

Work package 3:	Earth system observations
Deliverable D3.1:	Data catalogue
Type:	Report
Author:	Alex Stickler (UBERN)
Reviewers:	Leo Haimberger (UNIVIE), Hans Hersbach (ECMWF)
Delivered:	11 June 2014
Notes:	This report combines Deliverable 3.1 and 3.2



Grant agreement no. 607029



ERA-CLIM2: WP3 report on deliverables D3.1 and D3.2

A. Stickler

D3.1: A catalogue of additional historical meteorological data available as hard copy, imaged documents or microfilm, D3.2: A prioritised list of datasets to be rescued based on D3.1

These deliverables essentially consist of the searchable ERA-CLIM metadatabase that has been developed in the framework of the first part of the project, and that is freely accessible online under <http://www.oeschger-data.unibe.ch/metads/>. WP3 members are regularly updating the metainformation contained in this database. Updating can be done either manually for each record by all WP members, or by uploading the whole inventory for any type of data (upper-air, moving upper-air, surface, atmospheric transmission) by UBERN. In the following, a short summary of the status of the data catalogue and of the prioritisation for digitisation is given for each WP3 member, plus a brief, tabular overview of the largest data sources to be digitised in the framework of ERA-CLIM2.

UBERN

As detailed in the ERA-CLIM2 Kickoff Meeting in Reading on 24 and 25 February, the highest priority for UBERN is the completion of the digitisation of important upper-air data sources whose digitisation had already started during ERA-CLIM. The first three sources are volumes 1948 and 1951-53 of the *Daily Weather Report (Egypt)*, volume 1956 of the *Pakistan Daily Weather Report*, and volumes 1920-34 of the *Täglicher Wetterbericht* from Germany. The first two sources plus volumes 1920-26 of the third source have been downloaded from the NOAA Central Library and volumes 1927-34 of the third source imaged at UBERN during the first part of the project. The first source has meanwhile been completely digitised, whereas the second and the third are still being worked on.

A further high priority source are volumes 1931/32 and 1934 of the *Upper Air Data India*, which have been imaged at UBERN. In this case, volume 1931 is still being digitised, while volumes 1932 and 1934 have already been completed. Another high priority early source of upper-air data that is currently being digitised is volume 1903/04, the earliest part, of the Lindenberg Observatory record (apart the 1901/02 observations at its predecessor observatory at Tegel (Berlin), that have been digitised during ERA-CLIM). More high priority sources are

volumes 1926-40 of the aerological observations from the Netherlands, the International Days observations, volumes 1923 and 1925-28, and volumes 1922-25 and 1932/33 of the Meteorological Yearbook of Finland (all imaged at UBERN, but digitisation not yet started).

Some small, but high priority records containing data from expeditions and cruises to polar regions (observations on board the steam boat *Oihonna* 1901/02, measurements on board the Swedish survey vessel *Skagerak* 1906, Russina steam boat *Neptun* 1910, Norwegian Nowaja Semlja Expedition 1921, Greenland Expedition of the Univ. of Michigan 1929-31) have been finished or are being digitised at the moment. Further small, inventoried sources plus the results from observations on board manned balloons in Central Europe during the period 1888-1906 have been classified as medium (before 1900) or high priority.

Moving upper-air data are contained in some small sources, the International Days observations, the manned balloon data, and the *Täglicher Wetterbericht*. The complete metadata of the respective records are contained in the metadatabase. Table 1 gives a summary of the largest data sources to be digitised in the framework of ERA-CLIM2 at UBERN.

Table 1. Largest upper-air data sources to be digitised in ERA-CLIM2 at UBERN.

Source	Start year	Start month	End year	End month	Estimated station days	Observed parameters
<i>Täglicher Wetterbericht</i>	1920	1	1934	12	ca. 200,000	wind speed & direction, pressure, temperature, rel. humidity
<i>Upper Air Data India</i>	1931	1	1934	12	ca. 50,000-70,000	wind speed & direction, pressure, temperature, rel. humidity

Table 1. Continued

International Days	1923	1	1928	11	53,750	wind speed & direction, pressure, geopot. height, temperature, rel. humidity
Aerological observations (Netherlands)	1926	1	1940	12	ca. 30,000	wind speed & direction, pressure, temperature, rel. humidity
<i>Daily Weather Report (Egypt)</i>	1948	1	1953	12	ca. 10,000-15,000	wind speed & direction, pressure, temperature
<i>Pakistan Daily Weather Report</i>	1956	1	1956	12	ca. 13,000	wind speed & direction
Meteor. Yearbook Finland	1922	1	1933	12	ca. 10,000	wind speed & direction, pressure, temperature, rel. humidity

FFCUL

During ERA-CLIM2, FFCUL will be adding digitised data to the previous sets of surface and upper-air data that was supplied in ERA-CLIM. New publications were found in the IDL/FFCUL archives, relative to Angolan and Mozambican sub-daily surface data, for several stations already considered in the previous inventory and new ones. These comprise the periods 1937 to 1974 of the *Colónia de Angola – Elementos Meteorológicos e Climatológicos* logbooks and the periods 1934 to 1956 (with a gap between 1945 and 1950) of the *Província de Moçambique – Boletim mensal das observações meteorológicas feitas nas estações da Província*. From these we are giving high priority to the period 1947-1974

for Angolan stations and 1951-1956 period for the Mozambican stations. The Angolan publications have already been imaged and digitisation is completed for 1947-1958 and is expected to finish before the end of 2014. The Mozambican ones are in line for imaging, and subsequent digitisation during 2014 and 2015.

Meanwhile, as presented in the ERA-CLIM2 kick-off meeting in February this year, FFCUL is performing digitisation of surface Chilean data (43 stations) already imaged for 1950-1958 by the UPAC ERA-CLIM group. This work is now in progress, having been given high priority, with the year 1953 being recovered for about 8 stations. We expect to accelerate this process as soon as we are able to subcontract a person this year and have the 1950-1958 period finished by the end of 2015. Additional digitised data for the same Chilean stations has been supplied by the Meteorological Chilean Service, which is cooperating fully with FFCUL in relation to metadata information, for the period 1959-2012, with several gaps. We have given medium priority to the Quality Control and formatting processes of this dataset in relation to the 1950-1958 data.

Surface data recovered includes pressure, temperature, water vapour pressure, relative humidity, wind direction and speed, cloud cover, precipitation, evaporation, sunshine duration, grass temperature, deep soil temperatures and eventually cloud type information from the Chilean data.

In relation to the South China Sea 1900-1932 surface pressure datasets FFCUL has supplied to ERA-CLIM (in cooperation with ACRE and digitised by the UKMO group), we are digitising the temperature and wind data for the Macau station from 1894 to 1915 in order to complement the sub-daily data for this ex-Portuguese colony. High priority has also been given to the formatting of the subsequent period 1933-1941 of the South China Sea surface pressure already provided to FFCUL by the UKMO. The data source of this data are the Hong Kong and China logbooks on-line at the NOAA archive.

Upper air data (pilot balloon) for Spain starting in 1912 has been found in the IDL/FFCUL archives in the *Anuarios del Observatorio Central Meteorológico – Madrid* (1912-1915 Vol I-III) for 4 stations (Madrid, Barcelona, La Coruña and Las Cañadas del Teide-Tenerife). It was found that this dataset was important to digitise in order to fill the Iberian Peninsula region which has very sparse data before the 1930's. The Madrid station is already being digitised and imaging of the logbooks is being prepared. We will also digitise upper-air data from the Ebro (Catalonia) observatory for the 1930-1937 and 1942-1961 periods contained in the publications *Boletín mensual del Observatorio del Ebro* and *Boletín del Observatorio del*

Ebro – Meteorología, again belonging to the IDL library. Upper air data recovered will consist mainly of wind and altitude records from the pilot balloon data and eventually temperature, geopotential height and relative humidity from any radiosonde records. Tables 2 and 3 summarise the largest, high priority surface and upper-air data sources to be digitised in the framework of ERA-CLIM2 at FFCUL.

Table 2. Largest surface data sources with high priority to be digitised in ERA-CLIM2 at FFCUL.

Source	Start year	Start month	End year	End month	Estimated station days	Observed parameters
<i>Colónia de Angola – Elementos Meteorológicos e Climatológicos logbooks</i>	1947	1	1974	12	ca. 71,000	wind speed & direction, pressure, temperature, rel. humidity, water vapour pressure, cloud cover, precipitation, evaporation, sunshine duration
<i>Província de Moçambique – Boletim mensal das observações meteorológicas feitas nas estações da Província</i>	1951	1	1956	12	ca. 14,000	wind speed & direction, pressure, temperature, rel. humidity, water vapour pressure, cloud cover, precipitation, evaporation, sunshine duration, grass temperature, deep soil temperatures
Servicio de Meteorologia Aeronáutica / Fuerza Aerea del Chile	1950	1	1958	12	65000	wind speed & direction, pressure, temperature, rel. humidity, dew point temperature, cloud cover, precipitation, cloud type information

Table 3. Largest upper-air data sources with high priority to be digitised in ERA-CLIM2 at FFCUL.

<i>Anuarios del Observatorio Central Meteorológico – Madrid (1912-1915 Vol I-III)</i>	1912	1	1915	12	ca. 4,000	wind speed & direction, altitude
<i>Boletín mensual del Observatorio del Ebro and Boletín del Observatorio del Ebro – Meteorología</i>	1930	1	1961	12	ca. 10,000	wind speed & direction, altitude (and/or pressure, temperature and relative humidity)

RIHMI

The first part of the digitisation work at RIHMI encompasses the rescue of upper-air data. This is given the highest priority. A group of upper-air stations (38 stations from Russian territory) have been identified. A review has shown that the time period covered is varying from station to station, the earliest is starting in 1936, the latest ending in December 1960. The data are disseminated over three sources:

1. computer media (not all stations, poor vertical resolution - not all levels, few standard pressure levels only)
2. handwritten tables
3. digitised forms of old punch-card formats

An analysis has shown that sources 2 and 3 are much bigger than source 1. Merging all 3 sources to proper “synthesised” soundings allows to get a consistent and complete dataset with a maximal temporal and vertical resolution. The resulting data will be much more complete than any known digitised dataset of these stations (as e.g. in IGRA). The variables contained are geopotential height, air pressure, temperature, humidity parameters and wind parameters (speed and direction).

Current tasks are

- creating an inventory of these data and, after checking the station names and coordinates, putting it into agreed format
- digitising handwritten parts of data (ongoing)
- software programming for data reformatting, data merging, statistics calculation and plotting (in preparation)

The first test completed record (for one station) will be ready to be sent to ECMWF and UBERN by the end of June 2014.

The second part of the work is the digitisation of terrestrial data (e.g. snow cover) and sub-daily meteorological station data, some as far back as 1891 (medium priority). These data consists of standard sets of surface meteorological observations, snow depth and snow water equivalent for stations, and snow course observations.

Current tasks are

- assessing the sources of sub-daily data for an extension to the years before 1966
- preparing an inventory of the data to be digitised
- programming the software for merging the data from computer media and the newly digitised data, reformatting and checking the final data

The third part considers marine meteorological data that is not contained in ICOADS with low priority.

Current tasks are:

- assessing sources of data
- preparing an inventory of the data to be digitised

The fourth part is the digitisation of additional, „exotic“ sources of atmospheric (upper-air) data (low priority). This refers to episodic observations of the atmosphere in the early 1900s,

as kite observations, etc., that will be assessed and possibly digitised. This effort needs a case-by-case approach, as soon as formats of hardcopy sources differ essentially. The volume of data is essentially less than in the first item.

Table 4. Main sources of data to be digitised and prepared as datasets at RIHMI.

Source	Start year	Start month	End year	End month	Estimated station days	Observed parameters	Comments
Handwritten tables, computerised views of old punch-card formats, separate soundings on computer media (standard pressure levels only)	1936	1	1960	12	ca. 500,000	wind speed & wind direction, temperature, pressure, geopotential height, relative/specific humidity	For most stations, the start year is in the second half of the 1940s, each day has 3 or 4 soundings

METFR

As presented in the ERA-CLIM2 kickoff meeting, the highest priority for METFR is the digitisation of important upper-air data sources whose inventorying and imaging had already been carried out during ERA-CLIM.

During ERA-CLIM2, METFR will give the highest priority to the French overseas territories (Martinique, Guadeloupe, French Guiana, St Pierre et Miquelon, New Caledonia, French Polynesia, Réunion Island) upper-air data for the period 1938-1957, and to the French Southern Austral and Antarctic Territories (New Amsterdam, Kerguelen and Adélie Land) upper-air data for the period 1950-1957.

The first data source for French upper-wind data from pilot balloons before 1957 is the collection of original handwritten 4-pages daily reports *Compte Rendu Quotidien* containing

Table 4. Continued.

Handwritten tables, computerised views of old punch-card formats	1891	1	1965	12	300,000 – 400,000	Mean air temperature, minimum air temperature, maximum air temperature, horizontal visibility, total cloudiness, wind direction, mean wind speed, maximum wind speed, precipitation, soil surface temperature, water vapour partial pressure, relative air humidity, dew point temperature, air pressure at station level, air pressure at sea level.	Each day has 3 or 4 observations
Handwritten tables	1900 as the earliest	1	1936	12	c.a. 100,000	Snow depth	

upper-wind data on the last page: speed and direction given on pre-printed geometrical altitude levels. The collection of this type of document is to date the unique source for long-term pilot balloon series for the entire mainland France during the period 1923-1936, for the overseas departments during the period 1938-1957, and for the French Southern and Antarctic Territories (FSAT) during the period 1950-1957. The data source has been partially digitised during ERA-CLIM, mainly for France mainland. Most of the daily reports for the French territories have already been imaged during ERA-CLIM. The Réunion Island daily reports are in line for imaging and subsequent digitisation in 2015. METFR is performing digitisation of upper-wind for Martinique, Guadeloupe, New Caledonia and Kerguelen that is expected to be finished at the end of 2014. It may also be noted that several inventoried records from these daily reports for France mainland (1923-1948) have been classified as medium priority because of the very high density of upper-air stations in France mainland at that time. High priority is given to 15 long-term series including the Corsica Island series (Ajaccio, Angoulême, Antibes, Argentan, Bordeaux, Brest, Le Bourget, Istres, Lyon, Nancy, Nice, Nîmes, Saint-Inglevert, Strasbourg, Toulouse). These long-term series have been partially imaged and digitised during ERA-CLIM and the completion of the digitisation of these series is high priority.

Furthermore, high priority has also been given to the French radiosonde data for the period 1937-1957. The source for the first operational radiosonde observations in mainland France 1937-1939 are the *Observations Quotidiennes*, an institutional publication containing the early operational radiosonde data for Bordeaux, Chartres and Trappes stations, that will be digitised. The largest source for French radiosonde data for the period 1945-1957 is the collection of aerological reports. These reports, archived in several Météo-France centres, have been partially imaged during ERA-CLIM. Digitisation has not yet started. Imaging is now in progress and the start of the digitisation is expected in 2015. The digitised data will complete the previous sets of radiosonde data that was supplied in ERA-CLIM. Parameters to be recovered include wind, temperature, pressure and relative humidity.

A further high priority source is the *Journal officiel des Etablissements Français de l'Océanie* containing Polynesian pilot balloon data for the period 1938-1957, already imaged at METFR during ERA-CLIM. Digitisation is now in progress. Further, originals archives with pilot balloons data from Réunion Island (Saint-Denis station) are being imaged during 2014 for the period 1953-1957. The preparation of digitisation is expected in 2015.

Furthermore, high priority has been given to the International Days observations for the period 1900-1912. Three stations from France mainland are concerned: Trappes, Strasbourg and Itteville. All the volumes have been imaged at METFR during ERA-CLIM but digitisation has not yet started. Digitisation is expected to be prepared in 2015.

We have given medium priority to the North African bulletins (*Bulletin Quotidien d'Afrique du Nord*, period 1929-1930, 1933-1936, 1951-1957), containing North African upper-air data, because this collection is unfortunately incomplete and not easily readable. All these daily bulletins have already been imaged at METFR during ERA-CLIM.

Table 5 gives a summary of the largest, high priority upper-air data sources to be digitised in the framework of ERA-CLIM2 at METFR.

Some further small, inventoried sources from observations during the period 1898-1957 have been classified as medium priority.

Table 5. Largest upper-air data sources with high priority to be digitised in ERA-CLIM2 at METFR.

Source	Start year	End Year	Domain	Estimated station days	Observed parameters
<i>Compte Rendu Quotidien</i>	1938	1957	France overseas (12 stations)	ca. 30,000	wind speed & wind direction
<i>Compte Rendu Quotidien</i>	1950	1957	French Southern Austral and Antarctic Territories	ca. 3, 000	wind speed & wind direction
<i>Compte Rendu Quotidien</i>	1923	1948	France mainland (15 stations)	ca. 100,000	wind speed & wind direction
<i>Observations quotidiennes</i>	1937	1939	France mainland (3 stations)	ca. 3,000	wind speed, wind direction, temperature, pressure and relative humidity

Table 5. Continued.

<i>Compte Rendu Aérologique</i>	1945	1957	France mainland (7 stations)	ca. 26,000	wind speed, wind direction, temperature, pressure and relative humidity
<i>Compte Rendu Aérologique</i>	1951	1957	France overseas and Southern Austral and Antarctic Territories	ca. 7,000	wind speed, wind direction, temperature, pressure and relative humidity
Polynesian pilot balloon sheets	1938	1957	Polynesia (4 stations)	ca. 10,000	wind speed & wind direction
Réunion island pilot balloon sheets	1953	1957	Réunion island (1 station)	ca. 2,000	wind speed & wind direction
International Days	1900	1912	World (France mainland: 3 stations)	ca. 9,500	wind speed & wind direction, pressure, geopotial height, temperature and relative humidity