Data rescue and digitization: tips and tricks resulting from the Dutch experience

Theo Brandsma, Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands, theo.brandsma@knmi.nl

1. Introduction

From 1981 till about 1987 KNMI digitized part of the huge amount of pre-1850 instrumental meteorological data in the Netherlands. The activity was partly financed by the European Union. In total 0.4 million sub-daily observations[1] were digitized. Only recently the data were made freely available to the public at the KNMI website (http://www.knmi.nl/klimatologie). In the year 2000, KNMI renewed its efforts in the area of data rescue and digitization with a long-term activity (partly externally funded). Since then many types of data have been digitized and made available to the public. Examples of the data types that have been dealt with are: 18th and 19th century ship logs (http://www.knmi.nl/cliwoc/) amounting to 0.29 million observations, 19th century KNMI year books amounting to 0.6 million observations, films with observer log books of the Amsterdam City Water Office amounting to 1.6 million observations and logbooks with rainfall measurements in the 1850-1950 period amounting to 4.7 million observations. At present we are digitizing strip charts and paper rolls from pluviographs (321 stations years in total), the remainder of pre-1850 weather observations, data from the former colonies and metadata archives.

In this paper we discuss our experiences with digitization. We start by presenting examples of the materials that have been used as basis for digitization. Thereafter, we introduce three scanners that are being used for scanning the data, followed by a presentations of four methods used for getting the data into spreadsheets. We conclude the paper by a presentation of twelve do’s and don’ts.

2. Base material

2.1 Hardcopy (original or copy)

In many cases, especially in the past, observation books or logbooks were directly used as the source for keying in the data. If for some reasons these documents could not directly be used for digitization, paper copies of them were also being used. During the 1981-1987 digitization project at KNMI, we only used hardcopy data as source for digitization.

Consider some examples of hardcopy material that are being used in our projects. Figure 1 shows an example of handwritten observations from a 17th century ship log book. Figure 2 gives an example of tables with printed data from the 19th century KNMI yearbooks. As a last example, Figure 3 shows a rainfall strip chart with graphical information. The type of hardcopy material determines to a large extent the digitization method.

An important advantage of using original hardcopy base material is its readability. Disadvantages are the detoriation of the material during the digitization process and the fact that at the same time only one digitizer can work on the data (without making extra

[1] An observation is defined as the measurement of all available climate variables at the certain time and location.

copies of the data). Furthermore, the location for digitization is fixed to the location of the hardcopy data.

2.2 Films

Hardcopy historical climate data is sometimes stored in archives that are not willing to lend the data for digitization. They prefer making films of the data rather than making hardcopies. Copies of the films can mostly be requested or the films can be studied on the spot using a film-reader. As an example consider the hourly meteorological data (1784-1963) of the Amsterdam City Water Office. These data are stored in thousands of logbooks in the archives of the municipality of Amsterdam. The logbooks are, however, not allowed to leave the archive. In 1984 the archive put al the data on film and KNMI obtained a copy of them (Figure 4). KNMI has a film-reader that can be used to both view and print the data.

The quality of the images on films is mostly excellent and if high quality films are used (e.g. polyacetate films), they may last for more than 200 years. Therefore, films are an ideal source for preserving data. As with the hardcopy data, only one digitizer can work on a film (without making extra copies of them) and the location for digitization is fixed to the location of the films and the film-reader.

2.3 Digital images

Digital images of the data may be obtained by scanning or digitally photographing the hardcopy documents. There are also scanners available that can make digital images of films (or microfiches). In fact, the films of the Amsterdam City Water Office were transformed to digital images before keying the data into spreadsheets. For the actual keying of the data we used several working places with two screens next to each other, one with an image and one with a spreadsheet. With the increasing quality of scans and the growing storage capacity of computer systems the use of digital images as data source becomes more and more feasible.

The use of digital images has two important advantages compared the use of hardcopy data or films. First, the images can be used by more persons at the same time and at any location. Second, together with the digitized data also the images of the original data can easily be provided (e.g. via the Internet). An advantage of the latter is that the user has the possibility to go back to the original data. Especially for the older handwritten data this may be advantageous. Both digital images and films may serve as an extra backup of the hardcopy data in case of calamities. A disadvantage of digital images is the sustainability of the files. File formats like jpeg of giff will likely change in the future to other formats. This may require file conversions of the original files. In that process errors may easily be made.

3. Scanners

3.1 Bookscanner

Figure 5 shows one of our students working with the so-called CopiBook bookscanner (http://www.iiri.com/i2s/copibook.htm). We use this scanner for scanning books and old documents containing climate observations and metadata. The scanner has a book cradle to prevent damage of books and scans two color or grayscale pages with a resolution of 300 dpi in 7 or 2½ seconds, respectively. It provides, among others, automatic location and cropping of the documents. The new price of this type of scanner is about EUR 25.000,-. A much cheaper solution may be obtained by using a digital camera on a stand.

3.2 Large-format scanner

For the scanning of long paper rolls with registrations of self-recording rain gauges we obtained the Contex Chameleon G600 large-format scanner (http://www.contex.com). The rolls have a length of about 10 m and we were not allowed to physically cut the them into smaller pieces. The Contex large format scanners are one the very few scanners that allow for the scanning of long documents (limited only by the storage capacity of the computer). In Figure 6 the scanner is being used at KNMI for the scanning of the paper rolls. The Chameleon scans the rolls in color with a resolution of 400 dpi and a speed of 2.5 cm/s. It can handle documents with a width up to 1 m and we therefore use it also for digitizing maps. The new price of this type of scanner is about EUR 12.000,-.

3.3 Fast document scanner

We obtained the Canon DR5010C fast document scanner (http://www.canon-europe.com) for scanning about 100.000 strip chart self-recording rain gauges. In addition the scanner can also be used to quickly scan books whose cover and back may be removed. The scanner scans in color with a maximum resolution of 600 dpi and a speed of 20 strip charts or pages per minute. Figure 7 shows how the scanner is being used for scanning strip charts.

4. How to get the data into spreadsheets?

4.1 Manually typing

The most common method for getting data into spreadsheets is manually keying the data. For handwritten material this is still the only feasible method, but also for printed data it is often the most obvious method. The person that keys the data into the spreadsheet should learn to blindly use the numeric keypad (on the right hand side of the keyboard) with sufficient speed. The use of predefined shortcuts for entering non-numeric data may speed up the process. At KNMI we use manually typing for about 70% of our data.

4.2 Using Optical Character Recognition (OCR)

The use of OCR may be feasible when the observations are printed with sufficient quality in the documents. About 10 years ago KNMI experimented with this type of data entry for some of the KNMI yearbooks. The basis for OCR is the digital images of the data. When the quality of the originals and the images is poor, OCR may require a lot of post processing and, therefore, may probably not be much faster then manually typing. It should also be realized that the OCR results must be combined in files and quality checked. Since our experiment with the OCR software, the software may have improved and may be more suited to digitize printed climate records in old documents. We recently started experimenting with the OCR software ABBYY FineReader (http://www.abbyy.com/). In January 2008, a 4-year EU-project starts, focusing on the improvements of OCR for old printed documents (http://www.impact-project.eu). Note that OCR for handwritten material is still not feasible.

4.3 Using Speech Recognition Software(SRS)

In the year 2002 we experiment with the use of SRS. At that time, we were manually keying the 1.6 million observations of the Amsterdam City Water Office. It was hoped that the use of SRS would alleviate the manual typing of the data. We used the software named *Dragon NaturallySpeaking.* The software needs to learn the voice of the speaker and the speaker should be able to work without background noise. We found that the combination of the software with spreadsheets was not optimal and soon decided to go on with manually keying the data. However, it may be that newer SRS versions may be more feasible than the one used by us. In addition, when funds are available, it may be interesting to work with a specialized company to suite the SRS to you digitization needs.

4.4 Using automatic curve extraction software
Self-recording rain gauges have been applied for continuous rainfall measurements at a selected set of KNMI stations since the end of the 19th century. At first, rainfall was recorded on daily (Figure 3) and sometimes weekly rainfall strip charts. Thereafter, from about 1980 through 1993, paper rolls were used to register rainfall for about 10-20 days per roll. From 1994 onwards, rainfall measurements are transferred electronically and operationally stored at 10-minutes resolution (for some selected stations at 1-minute resolution). Until now, the strip charts and paper rolls have been used mainly for extracting hourly values. In infrastructural design (e.g. sewer systems, tunnel drainage) there is, however, a need for long rainfall series with much higher resolution than 1 hour. Fortunately, the charts and rolls can be used to extract rainfall with a time resolution of about 5 to 10 minutes.

We are developing a procedure that largely automates the labor-intensive extraction work for rainfall strip charts and paper rolls. Although developed for rainfall, it can be applied to other elements as well. The procedure consists of four basic steps: (1) scanning of the charts and rolls to high-resolution digital images using the scanners in Figures 6 and 7, (2) applying automatic curve extraction software in a batch process to determine the coordinates of cumulative rainfall lines on the images, (3) visually inspecting the results of the curve extraction, (4) post-processing of the curves that were not correctly determined in step (3). Although KNMI is still perfecting the software, several tens of station-years have succesfully been digitised. The time resolution is about 5 minutes. In total 321 station-years are being digitized. It is planned that the data will become available in 2009.

5. Do's and don'ts

5.1 Start with an inventory of historical climate data

Before starting to digitize, quality control and disclose historical climate data, an inventory should be made of all sources containing historical climate observations. Depending on the purpose the project, the inventory can be made on a scale ranging from a single institute (like a meteorological office) to a group of countries. The inventory should clearly show the opportunities for extending back in time the existing digital time series. It needs to reveal all known time series (both in hardcopy and digitized format), the station names, the observed parameters, their resolution, the observation period, the location where the data are stored, etc. In the year 2000 we made such an inventory at KNMI (Brandsma et al., 2000). An inventory may be helpful in the process of obtaining support and funds for the work that still needs to be done.

5.2 Check if the data is already available somewhere in digitized form

Unfortunately not all digitization efforts are well coordinated. As a result, it may not always be obvious if a data source has already been digitized. For example, consider a scientist that undertakes a digitization effort to digitize a particular time series needed for a scientific publication. After the series is digitized and used for the publication nobody cares about the series and after some time it may even be forgotten that the series has ever been digitized. The problem here is that to actually disclose the series to the public, some extra steps are needed that the scientist did not include in the planning of his digitization effort. For series of former colonies, it may be profitable to check the existence of digitized data internationally.

5.3 A professional should check the data source before having it digitized

Before starting the digitization, a professional should inspect the data source to look for changes in parameters, units, formats, missing data and other important metadata. This information is needed for constructing templates for keying the data into spreadsheets and for the construction of a metadata file. The templates should closely resemble the format of the data source to minimize keying errors. Sometimes summary measures are also provided in the data source. In that case it is advised to integrate formula in the spreadsheets to automatically calculate these measures from the keyed data. A comparison of the calculated summery measures with those in the data source may reveal keying errors. The amount of work required to construct digitizing templates, largely depends on the type of data source.

5.4 Gather all relevant metadata and supply them (at least in English) together with the digitized data

All relevant metadata concerning the series should be gathered and supplied together with the digitized data. Metadata may be found in the data sources themselves or separate publications. In addition, meteorological institutes often keep hardcopy files containing information about a particular station and the observed variables. The search for metadata may be time consuming.

It may be handy to distinguish between two types of metadata, which we define here as type I and type II metadata.
Type I metadata. Metadata needed to trace a times series. This type of metadata provides information about: location of the observations, time period(s) of the observations, observed variables, observation frequency and information about how the data are available (digital, hardcopy, quality of the data).
Type II metadata. Metadata needed (along with type I-metadata) to homogenize time series. This type of metadata provides information about: changes in instruments, relocation (horizontal and vertical) of the instrument, changes in methodology of the measurements, availability of parallel measurements and changes in the environment (growing of trees, urbanization, etc.).

For historical climate data it is common to enclose the most important metadata in the header of the file (often type I-metadata). Figure 8 gives an example of the header of one of the historical data files available via the KNMI website. Other metadata, like photographs, detailed descriptions of the data, analysis of the data, may be provided in separated reports or papers (often the type II-metadata).

5.5 Use university students for typing in your data

KNMI has good experience in working with students of the nearby University of Utrecht. The students get a part-time contract for the duration of the project, mostly for 4-12 hours a week. They are relatively free to plan their work according to their (ever changing) study schedule and the work pays better than many other student jobs. We request from the students that they can work fast with a good precision. We advise them to work not longer than 4-5 hours per day because most digitization work requires a level of concentration that cannot be kept the whole day.

It is recommended to experiment with the data to see what realistic digitization speed and accuracy of typing can be obtained by the digitizers. This helps to plan the allocation of manpower and serves as a guiding line for the digitizers. Their work should be checked regularly for speed and accuracy.

5.6 Don't forget to rigorously quality control your data

Quality control (QC) should be an integral part of each digitization project. QC procedures should correct for typing errors, accidental changes of columns, etc.

5.7 Don't assume that processing the data after digitization is routine work

Data are often keyed into several spreadsheets that need to be combined after finishing the actual digitization. Several errors can be made in that process. For instance, not all spreadsheets may be chronologically put together, or changes in columns or number of variables may not be adequately accounted for. These types of errors may me much more sincere than the occasional typing errors.

5.8 For many data sources it is worthwhile to supply images of the original

For the analysis and homogenization of historical climate data it is often needed to go back to the original data. The original data may contain important metadata that is not available elsewhere, like information about the changes in instruments and units. Also, when there has not been a rigorous QC it may be needed to check suspect values in the original data. For the latter reason KNMI decided to make scans of all data that was digitized in the 1981-1987 period. The scans will be provided together with the already available digitized data them via the KNMI website.

5.9 Use the operational infrastructure of your institute for centrally archiving (including back-up) of digitized data and digital images

It is important that the archiving of the digitized data and the corresponding images are integrated in the operational infrastructure of the institute. In this way all digital information can be centrally stored and backups can easily be made. At KNMI the historical climate data is available via the network drives and a backup is stored in a mass storage system. As an extra safety guarantee, a few times per year backups of the data are send to a supercomputer in Amsterdam.

5.10 Assess the sustainability of your storage media

Tapes, floppies, CD-Roms, films, etc. have all a limited life time. Moreover, also the hardware needed to read these storage media may slowly disappear. Therefore, each digitization project should consider the sustainability of the required storage media and the hardware needed to read them. If the data is stored and backups are made as described in 5.9, then the need for the mentioned storage media may diminish. In all cases, however, attention should be given to the sustainability of the file types. Common file types like pdf, jpeg, bmp, etc. are probably not everlasting and may need transformation in the future into other types.

5.11 Put the data freely on the internet and provide the world databases with a copy

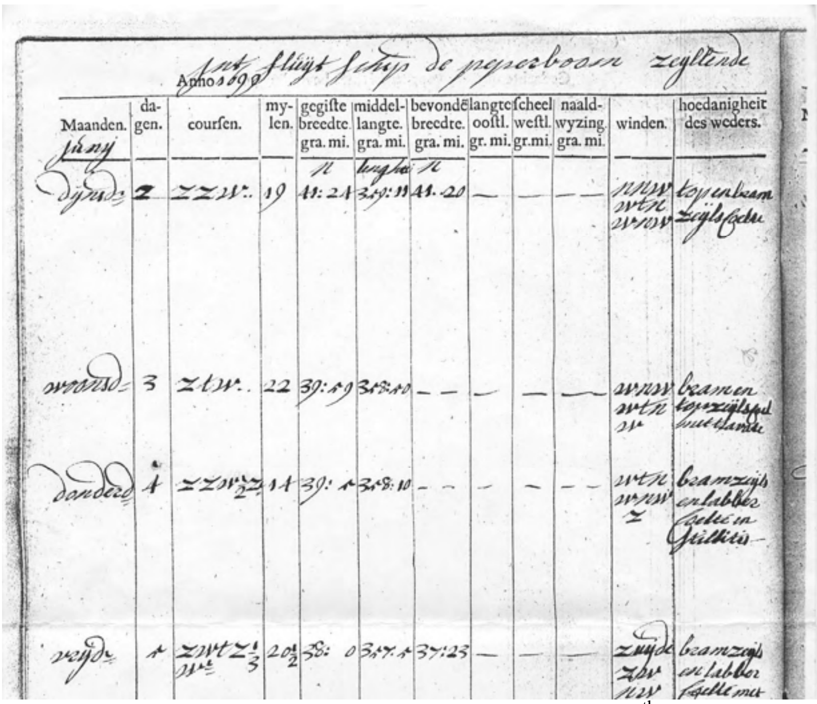
The commercial value of the majority of digital historical climate data is negligible. On the other hand, its scientific value cannot be underestimated. The power of many analyses for climate change and variability is in the existence of large datasets with high-quality historical climate data. These datasets exist because countries freely provide their historical data on their websites and/or to the world databases. As such, institutes that put their historical data freely on the Internet contribute to the study of climate change and variability. In addition, the free distribution of historical climate data may be a good publicity for the institute and (when people start using the data) may also help in detecting errors in the data and in the finding of new metadata.

5.12 Enjoy your work!!

Digitization and disclosure of historical climate data has a somewhat old-fashioned and dusty image. This is undeserved! The old climate data represent the reference that the World needs to assess the present and future climate. Numerous scientific journal papers and books are published each year making use of these data. Although the actual digitization work is often labor-intensive, new techniques are becoming available to scan, digitize and disclose the climate data via the Internet in a more efficient way than was possible previously. In summary, there is plenty of reason to enjoy this type of work.

References

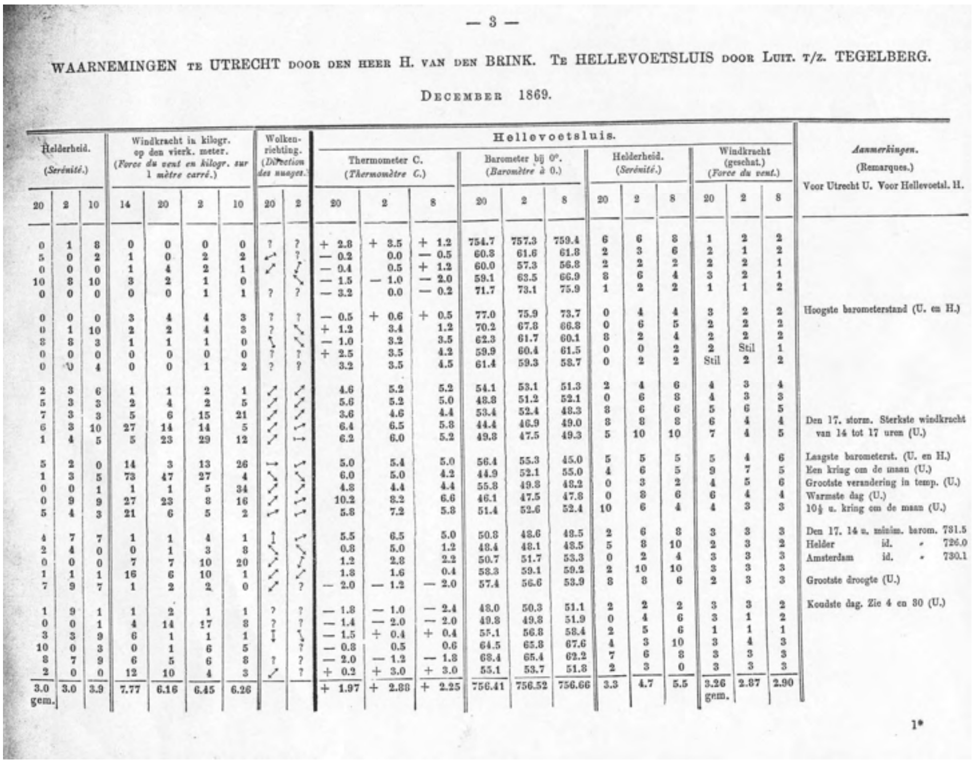
Brandsma, T., F. Koek, H. Wallbrink and G.P. Können, 2000. *Het KNMI-programma Hisklim (HIStorisch KLIMaat)*. KNMI-publication 191, also Hisklim 1, 72 pp.



Anno 1699 Ant fluyt schip de peperboom zeylende

Maanden.	da-gen.	courſen.	my-len.	gegiſte breedte. gra. mi.	middel-langte. gra. mi.	bevonde breedte. gra. mi.	langte ooſtl. gr. mi.	ſcheel weſtl. gr. mi.	naald-wyzing. gra. mi.	winden.	hoedanigheit des weders.
juny				N	lenghte	N					
dynsd	2	zzw.	19	41: 21	349: 11	41. 20	—	—	—	nnw wtn wnw	topzeilsen zeylsfoelte
woensd	3	ztw.	22	39: 59	348: 40	—	—	—	—	wnw wtn w	bramen topzeilsind hut hande
donderd	4	zzw	14	39: 1	348: 10	—	—	—	—	wtn wnw z	bramzeils en labber coelte en stiltien
vrydg	1	zwtz ½ w	20½	38: 0	347. 1	37:23	—	—	—	zuyde ztw nw	bramzeils en labber coelte met

Figure 1. Example of handwritten observations in a 17[th] century ship log book.



— 3 —

WAARNEMINGEN TE UTRECHT DOOR DEN HEER H. VAN DEN BRINK. TE HELLEVOETSLUIS DOOR LUIT. T/Z. TEGELBERG.

DECEMBER 1869.

Helderheid. (Sérénité.)			Windkracht in kilogr. op den vierk. meter. (Force du vent en kilogr. sur 1 mètre carré.)				Wolken-richting. (Direction des nuages.)		Hellevoetsluis. Thermometer C. (Thermomètre C.)			Barometer bij 0°. (Baromètre à 0.)			Helderheid. (Sérénité.)			Windkracht (geschat.) (Force du vent.)			Aanmerkingen. (Remarques.) Voor Utrecht U. Voor Hellevoetsl. H.
20	2	10	14	20	2	10	20	2	20	2	8	20	2	8	20	2	8	20	2	8	
0	1	8	0	0	0	0	?	?	+ 2.8	+ 3.5	+ 1.2	754.7	757.3	759.4	6	6	8	1	2	2	
5	0	2	1	0	2	2	↗	?	— 0.2	0.0	— 0.5	60.8	61.6	61.8	2	3	6	2	1	2	
0	0	0	1	4	2	1	↙	↗	— 0.4	0.5	+ 1.2	60.0	57.3	56.8	2	2	2	2	2	1	
10	8	10	3	2	1	0		↘	— 1.5	— 1.0	— 2.0	59.1	63.5	66.9	8	6	4	3	2	1	
0	0	0	0	0	1	1	?	?	— 3.2	0.0	— 0.2	71.7	73.1	75.9	1	2	2	1	1	2	
0	0	0	3	4	4	3	?	?	— 0.5	+ 0.6	+ 0.5	77.0	75.9	73.7	0	4	4	3	2	2	Hoogste barometerstand (U. en H.)
0	1	10	2	2	4	3	?	↘	+ 1.2	3.4	1.2	70.2	67.8	66.8	0	6	5	2	2	2	
8	8	3	1	1	1	0	↘	↘	— 1.0	3.2	3.5	62.3	61.7	60.1	8	2	4	2	2	2	
0	0	0	0	0	0	0	?	?	+ 2.5	3.5	4.2	59.9	60.4	61.5	0	0	2	2	Stil	1	
0	'U	4	0	0	1	2	?	?	3.2	3.5	4.5	61.4	59.3	58.7	0	2	2	Stil	2	2	
2	3	6	1	1	2	1	↙	↙	4.6	5.2	5.2	54.1	53.1	51.3	2	4	6	4	3	4	
5	3	3	2	4	2	5	↙	↙	5.6	5.2	5.0	48.8	51.2	52.1	0	6	8	4	3	3	
7	3	3	5	6	15	21	↙	↙	3.6	4.6	4.4	53.4	52.4	48.3	8	6	6	5	6	5	
6	3	10	27	14	14	5	↙	↗	6.4	6.5	5.8	44.4	46.9	49.0	8	8	8	6	4	4	Den 17. storm. Sterkste windkracht van 14 tot 17 uren (U.)
1	4	5	5	23	29	12	↙	→	6.2	6.0	5.2	49.8	47.5	49.3	5	10	10	7	4	5	
5	2	0	14	3	13	26	→	↗	5.0	5.4	5.0	56.4	55.3	45.0	5	5	5	5	4	6	Laagste barometerst. (U. en H.)
1	3	5	73	47	27	4	↘	↘	6.0	5.0	4.2	44.9	52.1	55.0	4	6	5	9	7	5	Een kring om de maan (U.)
0	0	1	1	1	5	34	↙	↙	4.8	4.4	4.4	55.8	49.8	48.2	0	3	2	4	5	6	Grootste verandering in temp. (U.)
0	9	9	27	23	8	16	↙	↗	10.2	8.2	6.6	46.1	47.5	47.8	0	8	6	6	4	4	Warmste dag (U.)
5	4	3	21	6	5	2	↗	↗	5.8	7.2	5.8	51.4	52.6	52.4	10	6	4	4	3	3	10½ u. kring om de maan (U.)
4	7	7	1	1	4	1	↕	↗	5.5	6.5	5.0	50.8	48.6	48.5	2	6	8	3	3	3	Den 17. 14 u. minim. barom. 731.5
2	4	0	0	1	3	8	↘	↘	0.8	5.0	1.2	48.4	48.1	48.5	5	8	10	2	3	2	Helder id. " 726.0
0	0	0	7	7	10	20	↙	↙	1.2	2.8	2.2	50.7	51.7	53.3	0	2	4	3	3	3	Amsterdam id. " 730.1
1	1	1	16	6	10	1	↙	↙	1.8	1.6	0.4	58.3	59.1	59.2	2	10	10	3	3	3	
7	9	7	1	2	2	0	↙	?	— 2.0	— 1.2	— 2.0	57.4	56.6	53.9	8	8	6	2	3	3	Grootste droogte (U.)
1	9	1	1	2	1	1	?	?	— 1.8	— 1.0	— 2.4	48.0	50.3	51.1	2	2	2	3	3	2	Koudste dag. Zie 4 en 30 (U.)
0	0	1	4	14	17	8	?	?	— 1.4	— 2.0	— 2.0	49.8	49.8	51.9	0	4	6	3	1	2	
3	3	9	6	1	1	1	↕	↘	— 1.5	+ 0.4	+ 0.4	55.1	56.8	58.4	2	5	6	1	1	1	
10	0	3	0	1	6	5		?	— 0.8	0.5	0.6	64.5	65.8	67.6	4	3	10	3	4	3	
8	7	9	6	5	6	8	?	?	— 2.0	— 1.2	— 1.8	68.4	65.4	62.2	7	6	8	3	3	3	
2	0	0	12	10	4	3	↙	?	+ 0.2	+ 3.0	+ 3.0	55.1	53.7	51.8	2	3	0	3	3	3	
3.0 gem.	3.0	3.9	7.77	6.16	6.45	6.26			+ 1.97	+ 2.88	+ 2.25	756.41	756.52	756.66	3.3	4.7	5.5	3.26 gem.	2.87	2.90	

1*

Figure 2. Example of printed data in the KNMI yearbook of 1869.

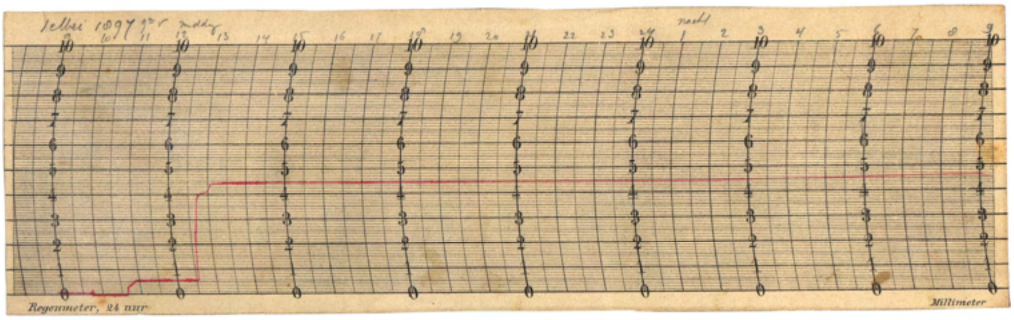


Figure 3. Example of a daily rainfall strip chart of De Bilt in 1897.


Figure 4. Copies of the films of the Amsterdam City Water Office in the archive of KNMI in De Bilt.

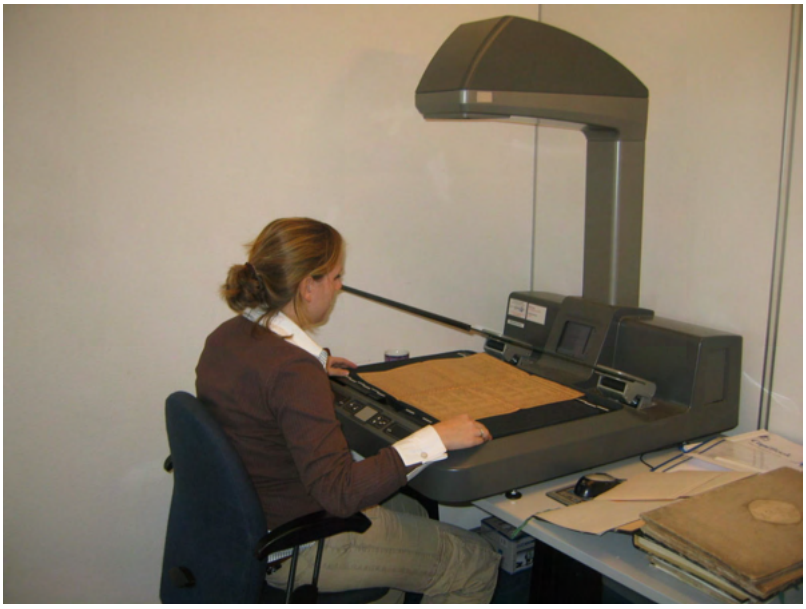


Figure 5. CopiBook bookscanner.


Figure 6. Contex Chameleon G600 large-format scanner.

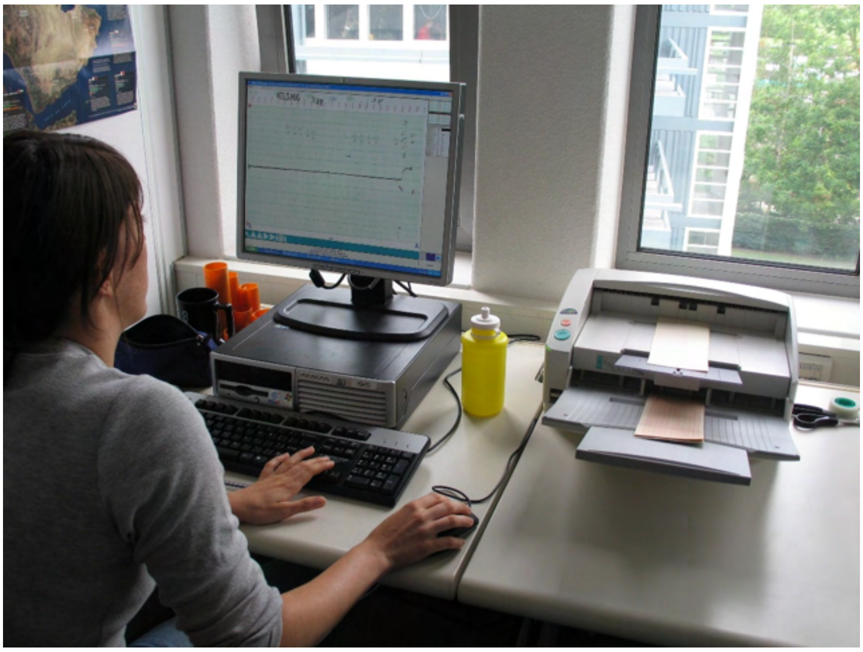


Figure 7. Canon DR5010C fast document scanner

```
senguerdius.dat - Kladblok
Bestand  Bewerken  Opmaak  Beeld  Help

DEZE GEGEVENS MOGEN VRIJ WORDEN GEBRUIKT MITS DE VOLGENDE BRONVERMELDING WORDT GEGEVEN:
KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT (KNMI)

THESE DATA CAN BE USED FREELY PROVIDED THAT THE FOLLOWING SOURCE IS ACKNOWLEDGED:
ROYAL NETHERLANDS METEOROLOGICAL INSTITUTE.

The data are obtained by digitalization of Senguerdius (1699) in the original units. Windspeed
and weattherterms are translated from Latin into Dutch.
This dataset contains the digitized and corrected data.

STN = LEIDEN,
YYYYMMDD = datum / date (YYYY=year MM=month DD=day)
PA, PB  = luchtdruk in oude Rijnlandse duimen en lijnen/ mercury stick barometer A and B / surface air
      pressure in old-style "Rijnlandse" inches and lines.
      first 2 digits is number of inches, next 2 numbers is number of lines.
      1 "Rijnlandse" inch(duim) = 12 lines; 1 Rhineland inch= 26.1518 mm; one line=2.18 mm
TA, TB, TC, TD  = Temperature A - D, in integer values of an unspecified unit. The unit differs
      among the four Thermometers. The units have in common that they all refer to an inverted
      scale (cold = high value).
DD  = windrichting / winddirection (N=North, O=East, Z=South, W=West)
FF  = windsnelheid in een 4 delige windschaal/ windspeed in a 4 point windscale (weinig wind;
      windstilte = low winds to windless, matige = moderate, krachtige = strong, stormachtige =stormy)
       WW  = luchtgesteldheid/ weather conditions only in Dutch
PA1, PB1 = surface air pressure of barometer A transformed into 0.1 mm Mercury.
TC1 = temperature of liquid thermometer C transformed into 0.1 degrees Celsius.
TB1 = temperature of air thermometer B transformed into 0.1 degrees Celsius.
PA2 = air pressure of barometer A in 0.1 hPa and reduced to 0 degrees Celsius with thermometer B.
PB2 = air pressure of barometer B in 0.1 hPa and reduced to 0 degrees Celsius with thermometer B.
PA3 = air pressure of barometer A in 0.1 hPa and reduced to 0 degrees Celsius with thermometer C.
PB3 = air pressure of barometer B in 0.1 hPa and reduced to 0 degrees Celsius with thermometer C.
PA4 = air pressure of barometer A in 0.1 hPa adjusted to the long-term modern day mean air pressure.

   STN  YYYYMMDD    PA    PB    TA    TB    TC    TD DD   FF                          WW

LEIDEN ,16970201, 2810, 2810,  114,  117,-9999,-9999,w   ,matige wind                ,wisselend bewolkt; rustig; buien met re
LEIDEN ,16970202, 2810, 2810,  111,  115,-9999,-9999,nnw ,matige wind                ,zonnige perioden en sneeuwbuien; afkoel
LEIDEN ,16970203, 2810, 2810,  118,  133,-9999,-9999,no  ,matige wind                ,'s ochtends bewolkt; 's middags zonnige
LEIDEN ,16970204, 2808, 2806,  131,  150,-9999,-9999,no  ,krachtige wind             ,zonnige perioden; strenge vorst
LEIDEN ,16970205, 2807, 2806,  134,  150,-9999,-9999,ono ,stormachtige wind          ,zonnige perioden; strenge vorst
LEIDEN ,16970206, 2807, 2808,  131,  140,-9999,-9999,ono ,stormachtige wind          ,betrokken; strenge vorst
LEIDEN ,16970207, 2809, 2808,  125,  134,-9999,-9999,no  ,matige wind                ,betrokken; vorst
LEIDEN ,16970208, 2808, 2808,  122,  128,-9999,-9999,z   ,matige wind                ,betrokken; vorst en sneeuw
LEIDEN ,16970209, 2806, 2806,  113,  112,-9999,-9999,wzw ,krachtige wind             ,zonnige perioden; opwarmend; 's middags
LEIDEN ,16970210, 2806, 2807,  120,  126,-9999,-9999,nw  ,weinig wind; windstilte    ,betrokken; van tijd tot tijd hagel; reg
LEIDEN ,16970211, 2808, 2808,  123,  133,-9999,-9999,nno ,matige wind                ,bewolkt; vorst
LEIDEN ,16970212, 2808, 2808,  123,  133,-9999,-9999,nno ,weinig wind; windstilte    ,bewolkt; vorst
LEIDEN ,16970213, 2810, 2809,  118,  126,-9999,-9999,ozo ,weinig wind; windstilte    ,bewolkt; vorst
LEIDEN ,16970214, 2807, 2806,  111,  112,-9999,-9999,zw  ,weinig wind; windstilte    ,bewolkt; vorst
LEIDEN ,16970215, 2806, 2806,  109,  110,-9999,-9999,zo  ,weinig wind; windstilte    ,bewolkt; vorst
LEIDEN ,16970216, 2800, 2801,   92,   94,-9999,-9999,no  ,weinig wind; windstilte    ,bewolkt; vorst; 's middags sneeuw
LEIDEN ,16970217, 2801, 2711,  100,   97,-9999,-9999,zw  ,weinig wind; windstilte    ,bewolkt; vorst en sneeuw
```

Figure 8. Example of the header the file with historical data from Leiden (1697-1698) as available form the KNMI-website at: http://www.knmi.nl/klimatologie/daggegevens/antieke_wrn/.