TASK ORDER # L-21-06

CLIMATE DATABASE MODERNIZATION PROGRAM

April 26, 2006

TASK ORDER TITLE:

KEYING THE URUGUAY SURFACE DAILY RECORDS

Task Leader:

Vin Stanton

Task objective:

The work involved with this task includes keying daily surface observation records from Colonia, Uruguay for October 1949 thru December 1972. This is a period of relatively consistent observational forms and was selected for the first keying phase. Earlier periods are available in varying formats and will be keyed later. The images will be provided to the contractor on CD-ROMs (each year will be on a separate CD arranged in monthly files) by NCDC. The keyed data shall be provided to NCDC via FTP in the prescribed output keying format (see Attachment A). The final keyed data will be placed on the FTP site in a sub-folder for Uruguay named Colonia. The keying format may be used again if more like data are received.

Pilot test / **Cost estimate:** The contractor shall do a pilot test using data from October-December 1949 (the first CD to be issued) to determine a keystroke estimate and approximate cost for this task. The cost estimate shall be submitted to CDMP prior to approval for production keying. All of the images for Colonia will be provided to the contractor so they can compute a keystroke count to complete the station. The total cost for the project shall be based on the total number of keystrokes in accordance with GSA schedule pricing. CDMP wants to ensure a competitive price with a fairly accurate budget.

Task details:

The contractor shall design and implement a production keying process for the Colonia, Uruguay Surface Observation Forms to ensure that all observational forms are keyed and output as specified. NCDC will identify a set of forms for test keying by the contractor. The test set will contain fairly representative samples of the daily forms likely to be encountered and will assist in developing the keying process and output files. Also, the test set will be used by NCDC to evaluate the process and establish final approval for full production.

Keying format:

The contractor shall provide an accuracy rate of 99.0% on all keystrokes entered. To meet the accuracy rate standards, the data are to be key verified (double-keyed in the blind). These accuracy rates shall apply to each keyed file (batch) processed and sent to NCDC. If a batch is failed by CDMP then the entire batch must be reprocessed.

Keying instructions are provided for each section of the Surface Observation form.
The contractor shall provide an output comma delimited text file (Attachment A) for each year's data. The filename shall be in the following format: "**Colonia1949**.txt", where "**Colonia"** is the station name and "**1949"** is the year of the data keyed.

A tilde (~) is to be inserted by the keyers whenever an entry cannot be determined due to poor legibility. This will indicate that there was an entry on the form but that the keyer could not determine the value. Every effort should be made to enter the value and to only enter a tilde as a last resort. Based on our initial review of the images a tilde should be an extremely rare entry.

The observation forms will contain anywhere from one to a maximum of four observations per day. The observations are contained on two separate pages; data from both pages will make up a complete keyed record. The images for Oct 1949 through Dec 1950 have both pages of the observation on one image while for the other years there is a separate image for each page. More details are contained in the attached keying format.

The observation forms have some peculiar value entries.
For example, the value "1" generally appears on the form as " $\big\backslash$ ". Other peculiarities found with the data are provided on the following page.

The following examples show some of the unique problems with the Uruguay data.

V V	ww	W	PPP	TT	Nh	CL	h	CM	CH	6	E	a	pp	7	RR	Te Te
96	61	2	11.4	18	8	7	4	x	x	6	1	2	n	7		
96	63	6	11.5	16	8	7	4	x	x	6	12		02	7	14	16
96	10	6	07.7	17	8	7	3	x	x	6	1 5		18	7		
97	02	2	07.9	18	8	7	4	x	x	6	1		08	7	17	18

These two images show what I call "lazy zeros" (00)', looking more like "11" that are connected by a line above. This is a very sporadic problem (correct entry = 00).

These images also show "X" as a valid entry. As stated in the keying format, these will be keyed "as is".

Also both of these examples show the different ways that the observers enter the numbers 1, 7 and to a lesser extent the number 9.

V V	ww	W	PPP	TT	Nh	CL	h	CM	CH	6	E	a	pp	7	RR	Te Te
97	02	0	14.7	20	0	0	9	0	0	6	0	3	n	7		
98	03	0	16.3	20	0	0	9	5	0	6	0	1	16	7	n	19
97	02	2	17.5	22	7	5	4	2	x	6	0	1	04	7		
97	02	2	19.4	21	5	4	6	0	0	6	0	4	26	7	00	23

Each CD-ROM provided by NCDC will contain images for one year's worth of station data. On the CD, the daily data images are contained sequentially in a corresponding monthly folder (generally 1 or 2 images per day of the month (e.g., a 31 day month would have either 31 or 62 images) - fewer if some days are missing).

An Example of the file structure on the CD-ROM is presented below:
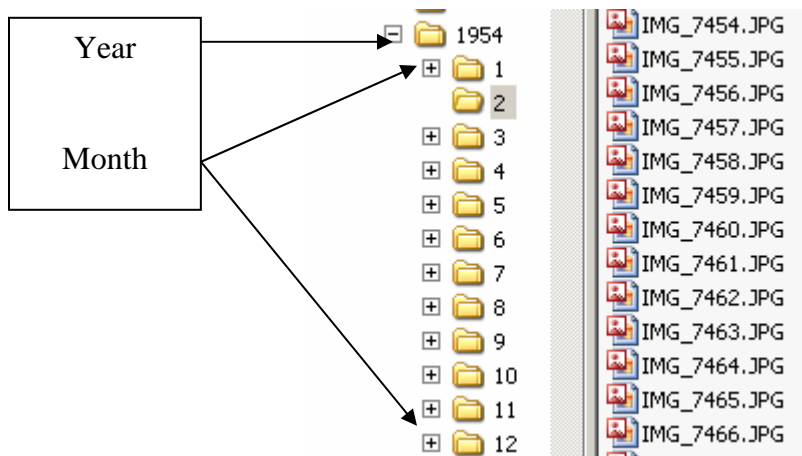
Specific details regarding the FTP transfer of keyed data shall be coordinated between the contractor and NCDC. The contractor shall maintain a backup copy of all keyed data (files) produced under this task until the project is completed. Upon project completion, NCDC will determine the disposition of the backup files.

If the contractor encounters any format inconsistencies, suspicious data or has any questions concerning the task, please contact NCDC. All recommendations and/or comments from the contractor for improving performance, efficiency and/or quality under this task are highly encouraged.

Attachment A
Output Keying Format for
Uruguay Surface Observation Forms

General Notes:

The contractor shall key hourly synoptic observations from Colonia, Uruguay, for the period
October 1949 through December 1972.   The images for Oct 1949-Dec 1950 contain page one
and page two on one image. For the remainder of the period, page one and page two are located
on two separate images.

For all entries, if the value to be keyed is not legible, the contractor shall key a single tilde ("~")
character (left-justified and blank-filled). Blank-fill any field if there's no available data to key
from the form. The number of available data entries on each form will vary from day to day.
When a field calls for two digits and only one is present, right justify and zero fill.
As always, any questions or clarifications should be directed to NCDC.

Keying Format for Uruguay "Surface Observation Form"

Position	Contents Column Header in Parentheses	Keying Instructions
1-6	WMO Station Number	865600, WMO # for Colonia
7	,	Comma delimited
8-11	Year	Four digit value (same for all images on the CD).
12	,	Comma delimited
13-14	Month	Two digit value, right justify and zero fill for values 1-9. Month of the subfolder on the CD.
15	,	Comma delimited
16-17	Day	Two digit value, right justify and zero fill for values 1-9. Day value found top left of page 1 (left page). The day is read from the image. The images are grouped in monthly files**.**
18	,	Comma delimited

19-22	TT (temp)	Temperatures are to tenths of degrees C
	(Tenths degree C)	decimal point is implied
	1st column	Key "-" for negative values in position 19

19-22   TT (temp)   Temperatures are to tenths of degrees C
          (Tenths degree C)   decimal point is implied
          1st column   Key "-" for negative values in position 19
Whole degrees in positions 20-21 (right justify, blank fill), tenths degree C in position 22, if missing blank fill. If no data present blank fill. If the character is unreadable insert a tilde (~).

NOTE: Data will be keyed from left most column and continue across the page for a total of four temperature (TT) entries. Times should agree with those entered in positions 68-69. For Period of record Oct 1949-Dec 1950, TT values are located on page 1 (left side page) row 15. From Jan 1951, TT values are located on page 1, row 13.

Data Entry Examples:

If form entry = -10.9
Then key – in position 19
       1 in position 20
       0 in position 21
       9 in position 22

If entry = 7.9, then key:
  Pos 19 = blank
  Pos 20 = blank
  Pos 21 = 7
  Pos 22 = 9

If entry = .9 then key:
Pos 19 = blank
Pos 20 = blank
Pos 21 = blank
Pos 22 = 9

23   ,   Comma delimited

24-27   TT (temp)
          (Tenths degree C)

	2nd column	Same as for positions 19-22
28	,	Comma delimited
29-32	TT (temp) (Tenths degree C) 3rd column	Same as for positions 19-22
33	,	Comma delimited
34-37	TT (temp) (Tenths degree C) 4th column	Same as for positions 19-22
38	,	Comma delimited
39-42	Wet Bulb Temp (Tenths degree C) 1st column	Follow the same instructions as for positions 19-22 **NOTE**: Data will be keyed from left most column and continue across the page for a total of four fields. For period of record Oct 1949-Dec 1950 the data is located on page 1 (left side page) row 16. From Jan 1951, the value is located on page 1, row 14.
43	,	Comma delimited
44-47	Wet Bulb Temp (Tenths degree C) 2nd column	Same as positions 39-42
48	,	Comma delimited
49-52	Wet Bulb Temp (Tenths degree C) 3rd column	same as positions 39-42
53	,	Comma delimited.
54-57	Wet Bulb Temp (Tenths degree C) 4th column	same as positions 39-42
58	,	Comma delimited.

59-61	Max Temp (T_xT_x) (Tenths degree C)	Positions 59-60 whole degrees C (right justify blank fill), position 61 tenths of degrees C, decimal point implied. This value is located on the right side (fourth column) page 1, row 21 from Jan 1951 – Dec 1972 and page 1, row 23 for the period Oct 1949-Dec 1950. If no data present, blank fill. Disregard any data contained in parenthesis next to the T_xT_x data entry.
62	,	Comma delimited.
63-66	Min Temp (T_mT_m) (Tenths degree C)	Position 63 reserved for negative sign (-) otherwise, blank fill. Positions 64-65 whole degrees C (right justify, blank fill). Position 66 tenths of degrees C, decimal implied. This value is located on left side (first column) page 1, row 22 for period Jan 1951 through Dec 1972. For period of record Oct 1949-Dec 1950, this data is located on page 1 left side (first column) row 24. If no data present, blank fill. Disregard any data contained in parenthesis next to the T_mT_m data entry.
67	,	Comma delimited.
68-69	Time (GG)	Right justify and zero fill if one digit reported for time. Data for positions 68 through 109 will be taken from row 47. **NOTE**: Data for the two pages on one image period, **October 1949-December 1950,** will be taken from row 50.
70	,	Comma delimited.
71	Total Cloud Cover (N)	Key as entered on form (valid entries = 0-9) **Note**: Column YY, iii, T_dT_d, 6, 7, and T_eT_e are not to be keyed on page 2.
72	,	Comma delimited.

73-74	Wind Direction (dd)	Key as entered, Right justify and zero fill. Valid entries range from 00-36 , 99 & XX
75	,	Comma delimited
76-77	Wind Speed (ff) (Knots)	Key as entered. Right justify and zero fill. Valid entries 00-99 & XX; notify NCDC if any entries >99
78	,	Comma delimited
79-80	Sfc Visibility (VV) (Meters)	Key as entered. valid entries 90-99
81	,	Comma delimited
82-83	Present Weather (ww)	Key as entered. Right justify, zero fill. Valid entries 00-99
84	,	Comma delimited
85	Weather past hour (W)	Key as entered. Valid entries 0-9
86	,	Comma delimited
87-89	Pressure (PPP) (tenths milibars)	Key as entered, decimal point is implied.
90	,	Comma delimited
91	Amount of low cloud (N_h)	Key as entered. Valid entries 0-9
92	,	Comma delimited
93	Type Low Cloud (C_l)	Key as entered. Valid entries 0-9 & X
94	,	Comma delimited
95	Height of lowest cloud (h)	Key as entered. Valid entries 0-9 & X
96	,	Comma delimited

97	Type Middle Cloud (C_m)	Key as entered. Valid entries 0-9 & X
98	,	Comma delimited
99	Type High Cloud (C_h)	Key as entered. Valid entries 0-9 & X
100	,	Comma delimited
101	E (State of Ground)	Key as entered. Valid entries 0-9
102	,	Comma delimited
103	Pressure Tendency (a)	Key as entered. Valid entries 0-8, Key any 9 entries and NCDC will perform the QC. Key as entered on the form (image).
104	,	Comma delimited
105-106	Pressure Change (pp)	Key as entered, right justify, zero fill.
107	,	Comma delimited
108-109	Precipitation in mm (RR)	Key as entered. Right justify, zero fill.
110	,	Comma delimited
111-153	Next hourly group	same instructions as for positions 68-110 **Note:** key row 51 images prior to 1951 and row 48 1951 and later
154-196	Next hourly group	same instructions as for positions 68-110 **Note:** key row 52 images prior to 1951 and row 49 1951 and later
197-238	Next hourly group	same instructions as for positions 68-109 **Note:** key row 53 images prior to 1951 and row 50 1951 and later (this has one less position since we do not want to end with a comma)