Status of NCDC Keying of Historical Marine Data

Joe D. Elms
National Climatic Data Center
Asheville, NC 28801

Abstract

U.S. historical marine weather observations (1912-1946) are currently being digitized at a rate of approximately 80 thousand per month at the National Climatic Data Center (NCDC). The existing quality control system, designed to help find (flag) keying errors, is being redesigned in an effort to improve efficiency so that the inspection (automated and manual) of the keyed records can keep pace with the digitizing rate. Approximately 2.5 million observations will be keyed under this project and should be available for the COADS Release 2 update. The history of the project along with some of the unique problems encountered while digitizing this historical marine data set are discussed.

Inventories based on the data assembled for the Comprehensive Ocean-Atmosphere Data Set (COADS, 1854-1979) Release 1 (Slutz et al., 1985) clearly indicate a lack of data during the World War I and II periods (reference Fig 1.). NOAA scientists, after conferring with COADS users, believed that digitizing additional marine data for these periods was warranted considering the significant contribution to the global climate from the world's oceans and the current concerns over climate change. In preparation for Release 2 (Elms,1991), an update of the COADS data base (Release 1 and Release 1a), a search of the National Archives for ship observations for these periods that were not in the existing digital data base was initiated. A significant number of World War I Merchant Marine observations were located (approximately 1.25 million); however, less than 5 thousand Merchant Marine weather observations were found for the World War II period (1942-1946). The Merchant Marine Logs dealing with personnel matters such as crew members aboard, sick call information, and seamen AWOL were located in the archives, but the corresponding weather logs were not found. It is theorized that the weather observations were probably declassified at the end of the war, returned to the parent shipping company and eventually destroyed. The search continues in an effort to verify the actual disposition of these valuable records.

The search revealed a number of Merchant Marine records (approximately 1.25 million) for the intervening period between the two World Wars (1918-1938) that were not in the COADS digital data base; therefore, these are also being digitized under the Release 2 initiative. The U.S. Merchant Marine observations for the period 1938-1948, including the fixed weather ship observations for 1940-1945, were maintained, at one time, on punched cards designated as Card Deck 115 (National Weather Records Center, 1953). These totaled approximately 946 thousand observations. In December 1959 it was decided by the "Card Deck Disposal Committee" that the number of errors existing in this particular deck warranted the cards disposal. In addition, the cards were not even filmed since the committee believed that many of the errors could not be corrected and it was likely that many of the observations were duplicated in other decks, particularly those acquired from foreign sources. In November 1960, the cards were destroyed

and because of this action it is now impossible to determine if any unique observations were lost. It is likely that a) a significant number of the observations were from the fixed ships which have been retained in other digital files, b) the foreign observations have been received through subsequent data exchanges and c) most of the U.S. observations are now being re-keyed under the current scheme; therefore, few observations should have been lost because of this decision.

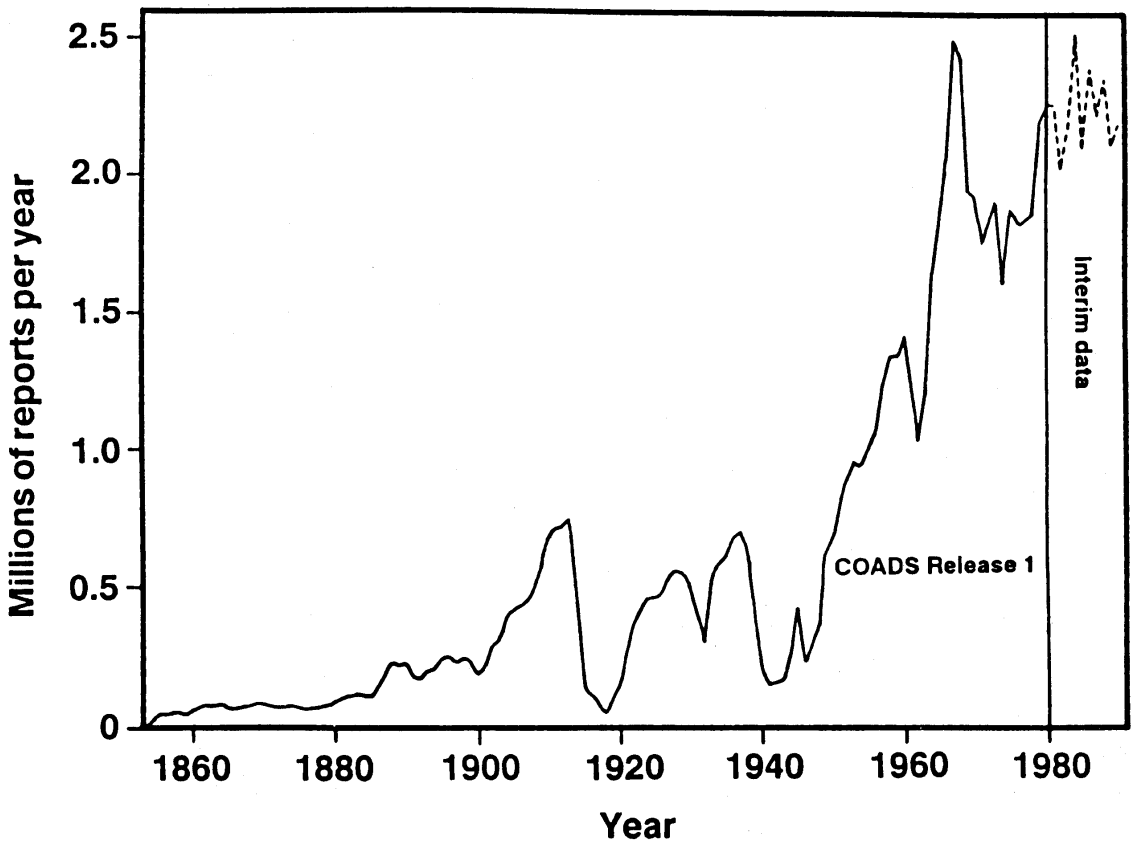


Figure 1. Number of global observations per year

In 1989, a pilot project began at NCDC to key the U.S. observations located in the National Archives that did not already reside in the digital data base for the period 1912-1946. It was the consensus at that time that the production keying to follow would likely be done outside the U.S.; therefore, CLICOM was selected as the system to be used for keying and quality controlling the data. A software development program was initiated (Bissinger and Manns, 1991) to adapt the CLICOM system originally designed for fixed land stations for keying mobile ship observations. In an effort to replicate as much detail as possible on the original ship forms, such as the units (e.g. Celsius, Kelvin, or Fahrenheit), over 20 different form types were created for the CLICOM keying effort.

Figure 2 illustrates three of the various form types. The largest segment of forms to be keyed contain only one observation per day (Fig. 2a), some had two per day (Fig. 2b), and beginning in

about 1938, four or more observations per day were generally recorded along with an increase in additional elements (Fig. 2c).

Figure 2. Example of ships' weather observation forms

Prior to June 1990, just over 40 thousand observations had been keyed on the CLICOM system under the pilot project and it was then decided to go into production using the CLICOM system. However, over the next few months it became clear that we would not be able to obtain the monthly production figures required to complete the keying by 1994. The CLICOM hardware

configuration prevented us from appending additional data entry ports; therefore, we decided to convert the keying process to the Viking system currently being used at NCDC for routine data entry. This was completed in July of 1991, and a second shift of keyers added because there were not sufficient data entry stations available on the day shift to handle the extra personnel needed to expand the marine keying effort. At the end of 1991 approximately three quarters of a million observations (Fig.3) had been digitized out of the approximate 2.5 million available. Figure 4 illustrates the expected increase by year in the number of additional observations from the U.S. holdings and from the Japanese Kobe Collection of Merchant Marine Ships (1890-1933). A significant number of additional observations (roughly estimated at 2-3 million) could be added if the Japanese Navy Data (1903-1944) were also digitized.

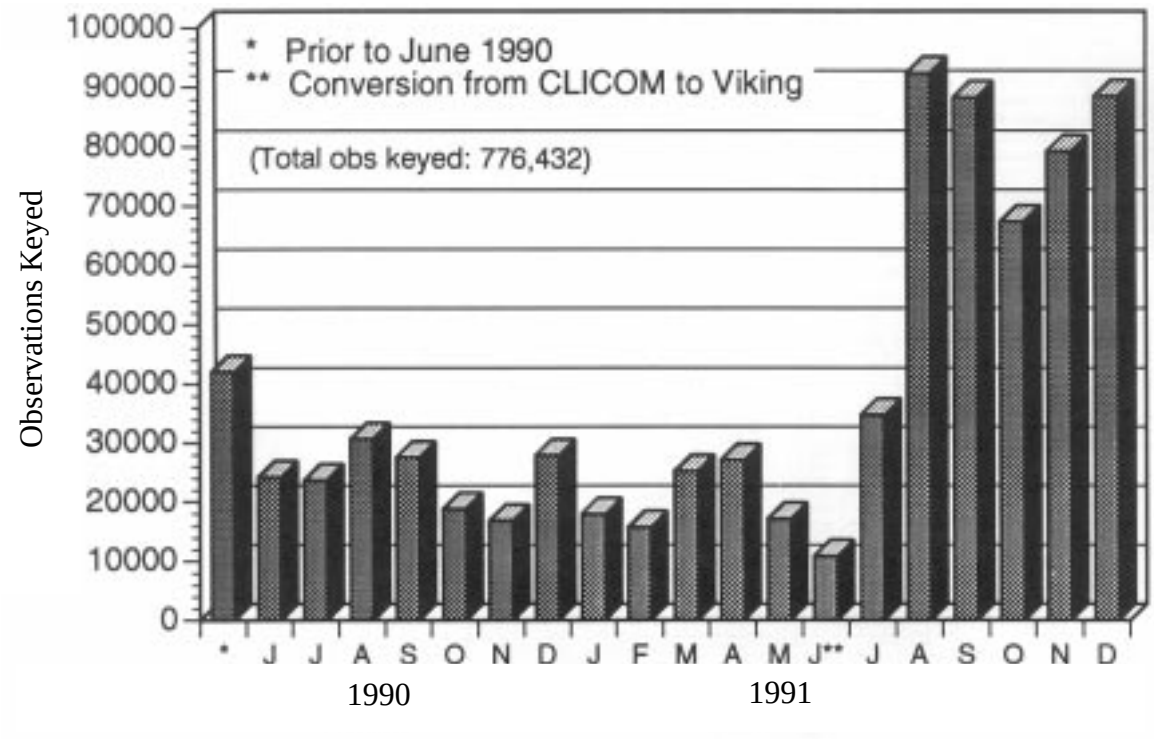


Figure 3. Number of observations keyed per month

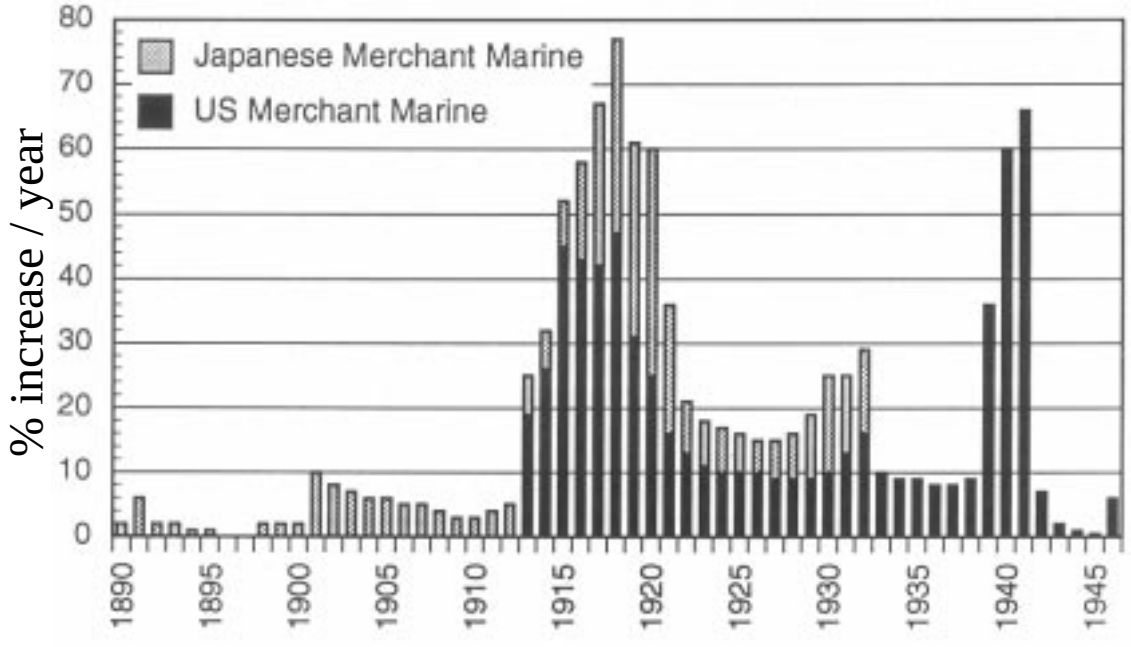
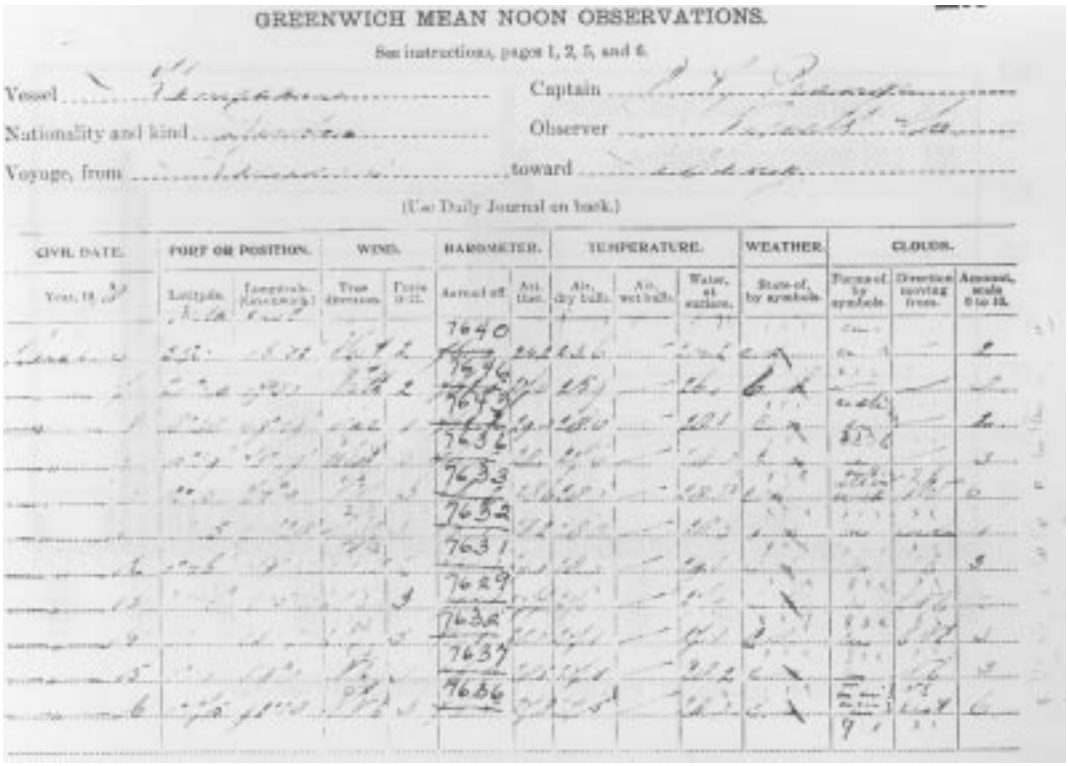


Figure 4. Expected increase in additional observations.

Numerous obstacles impede the digitizing effort such as the difficulty in interpreting the seamen's handwriting, (Fig. 5), and the deterioration of the forms caused by moisture and "bookworm" infestation (Fig. 6). Many data biases are introduced because of changing coding practices, lack of instructions to the observer, observers not adhering to the instructions, poor calibration of instrumentation, and insufficient metadata (information about the data). An example is the time bias introduced by recording only the Greenwich Mean Noon observation. At that time the U.S. Weather Bureau was interested in producing only one global chart per day, which they believed would be most useful in determining what type of weather conditions the mariner might encounter. In establishing this policy they obviously never considered the need for determining temperature trends and the need to investigate possible global warming!

In the early part of the 20th century, instructions to the observer requested that the wind direction be reported as the true direction and not the magnetic direction. However, nothing was specified as to whether it would be based on a 32 point scale or some other standard; therefore, while most reports were in the 32 point scale, the general practice at the time, many were coded at different scales such as 360 degrees, 16 or 36 point scales, or in one such as N76°E (meaning 76 degrees east of north). In our digitizing procedures, the wind directions are not always keyed directly as on the form, but under certain conditions converted to a 32 point scale. CLICOM also converts most alpha entries to a numeric doe for ease of handling.



GREENWICH MEAN NOON OBSERVATIONS.

See instructions, pages 1, 2, 5, and 6.

Vessel ..... Captain .....

Nationality and kind ..... Observer .....

Voyage, from ..... toward .....

(Use Daily Journal on back.)

CIVIL DATE.	PORT OR POSITION.		WIND.		BAROMETER.		TEMPERATURE.			WEATHER.	CLOUDS.		
Year, 18 ..	Latitude.	Longitude. (Greenwich.)	True direction.	Force 0-12.	Actual at	Att. (Inst.	Air, dry bulb.	Air, wet bulb.	Water, at surface.	State of, by symbols.	Forms of, by symbols.	Direction moving from.	Amount, scale 0 to 10.
					7640								
					7636								
					7633								
					7632								
					7631								
					7629								
					7628								
					7637								
					7636								

Figure 5. Example of seaman's handwriting

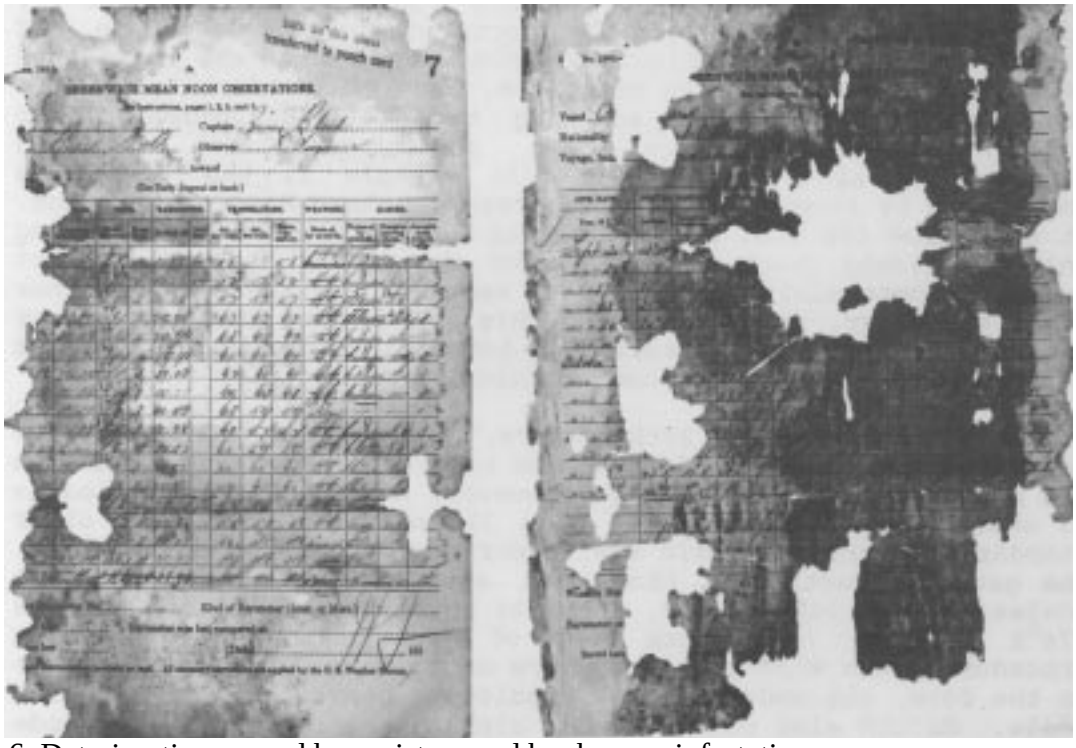


Figure 6. Deterioration caused by moisture and bookworm infestation

After it became common practice to transmit the weather observations over the ships radio, not only were more elements recorded, but both local and Greenwich Mean Times were entered on the observational form. Occasionally, these times are not compatible based on the reported ship's location. This requires human intervention to resolve these differences. Rarely, these differences cannot be resolved and therefore lead to a loss of data. Records are also lost whenever the ship position was not entered into the log (Fig. 7) or the date/time group contained an error that could not be rectified.

W. B. Form 1201—Marine, 1917.

SHEET A.

GREENWICH MEAN NOON OBSERVATIONS.

-08 5

Vessel U.S.S.RADNOR — Ocean Atlantic

Nationality and rig AMERICAN — Captain [illegible], Comm. Officer

Voyage from New York toward Brest, France — Observer [illegible], Navigator

DATE. Month.	Year.	TIME. Local mean time.	PORT OR POSITION. Latitude N. or S.	Longitude (Greenwich) E. or W.	WIND. True direction.	Force 1–12.	PRESSURE. Barometer as read.	Att. Ther.	Pressure corrected.	TEMPERATURE. Air, dry bulb.	Air, wet bulb.	Water at surface.	WEATHER. State of, by symbols.	CLOUDS. Form of, by symbols.	Direction moving from.	Amount, scale 0–10.	SEA. Direction of.	State of.
May	1919																	
4 – 8					ExS	4	30.08			53	52	52	BC	Cu	[illegible]S	V	V	S
8 – 12					ExN	3	30.06			54	51	52	[illegible]	Cu	[illegible]S	C		S
6		11:28	46°8 N	7°50 W	ExS	3	29.91			58	56	54	BC	SCu	[illegible]S	9		S
12 – 4					NE	4	30.00			51	51	51	BC	CuN	NE	6		S
4 – 8					NE	3	29.97			55	51	53	OC	SCu	NE	IO		S
8 – 12					ExS	3	29.93			58	56	54	BC	SCu	[illegible]S	9		S
12 – 4					NExE	2	29.92			61	56	53	BC	SCu	NE	8		S
4 – 8					E	3	29.84			54	53	53	OCR	CuN	NE	IO		S
8 – 12					ExN	3	29.80			55	54	54	OCR	N	[illegible]N	IO		S
9th		11:52			E	2	29.84			57	56	54	OC	CuN	SE	I		S
12 – 4					E	2	29.78			56	56	52	OC	CuN	SE	I		S
4 – 8					E	1	29.80			57	57	55	OC	CuN	SE	6		S
8 – 12					SE	2	29.84			57	56	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
12 – 4			Entered Gironde River															
11th			Departed			10			"									
8 – 12					W	2	30.03			55	55	52	BC	SCu	W	4		S
12 – 4					W	4	30.15			57	55	54	OC	SCu	W	3		S
4 – 8					SWxW	2	30.23			59	58	56	OC	SCu	W	3		S
8 – 12					W	2	30.24			56	58	56	BC	CS	W	6		S

Weather Bureau List Barometer No ..... Kind (Aneroid or Mercurial) ..... Barometer [illegible] last compared at .....

On (date) ..... Reads too high by (amt.) ..... been requested. Too low by (amt.) .....

See instructions on cover. Record barometer and thermometers precisely as read.

Figure 7. Sample log with missing ship positions

One element that will require detailed scrutiny before converting to a common format is the pressure reading. This is due partially to the evolution of instructions to the observer from the U.S.Weather Bureau (1910 & 1925). In the early 1900's they required the value entered on the form exactly as read off the barometer and attached thermometer. All the necessary corrections would be applied by the Weather Bureau after the report was received (corrections for temperature and gravity apply only to mercurial barometers). Once the observations were being routinely transmitted by radio, the Weather Bureau requested that the corrections be applied when transmitted and only then should the corrected value be entered on the form in the column labeled "corrected". In many instances these rules were not closely followed; therefore, all the corrected

values will require verification to ensure that they were computed correctly. This can be automated and performed at the time of conversion to a common format as long as all of the header information is available (type of barometer and to what extent it reads too high or low).

Much remains to be done under this project. The quality control portion on the CLICOM system is saturated with the addition of 80,000 observations per month and it has become very inefficient. Therefore, this function is to be converted over to a more efficient system such as the Local Area Network or Workstation. Plans are to start this development in the first quarter of 1992. In this revised system, fewer form types will be required since indicators will be incorporated into the design of the new system to define various units and variables. one major task ahead is to design software to convert all these various digitized data sets (form types) into a common format compatible with COADS.

In summary, plans are to have all the U.S. Merchant Marine observations (1912-1946) digitized by 1994 and ready for the major merge into the COADS data base as part of Release 2. As the project progresses, we plan to routinely update the metadata and then, in conjunction with Release 2, publish an updated reference manual. Plans are also being developed to have the remaining undigitized data from the Japanese Kobe (see the contribution by T. Uwai in this volume) and U.S. Maury Collections keyed. The Maury Collection was data originally collected by Matthew Fontaine Maury as Superintendent of the Depot of Charts and Instruments in Washington D.C. in the mid-1800's and now consists of ship observations basically between 1820 and 1860 with a few observations dating as early as the late 1790's to as late as the early 1900's. The original records are maintained in the National Archives, but have been copied on to 88 reels of microfilm for this project. Although it is unlikely the data from the Kobe or Maury Collections will be ready in time for Release 2, it should be available for the following update.

References

Bissinger, R. and Manns, D., 1991: *CLICOM - Flexible Data Management, Seventh International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology,* American Meteorological Society, Boston, MA, pp391396.

Elms, J.D., 1991: *Enhancement for the Global Marine Data Base, Seventh symposium on Meteorological observations and instrumentation,* American Meteorological Society, Boston, MA, J19-J23.

National Weather Records Center, 1953: Card Deck 115 Reference Manual, unpublished, available through the National Climatic Data Center.

Slutz, R.J., Lubker, S.J., Hiscox, J.D., Woodruff, S.D., Jenne, R.L., Joseph, D.H., Steurer, P.M., Elms, J.D., 1985: *COADS Comprehensive Ocean-Atmosphere Data Set Release 1*.

U.S. Weather Bureau, 1910: *Instructions to the Marine Meteorological Observer of the U.S. Weather Bureau,* Third Edition, W.B. No. 444.

U.S. Weather Bureau, 1925: *Instructions to Marine Meteorological Observers*, Circular M, Marine Division, Fourth Edition, W.B. No. 866.