

WORLD METEOROLOGICAL ORGANIZATION

INTERGOVERNMENTAL OCEANOGRAPHIC
COMMISSION (OF UNESCO)

JOINT WMO/IOC TECHNICAL COMMISSION FOR
OCEANOGRAPHY AND MARINE METEOROLOGY
(JCOMM)
EXPERT TEAM ON MARINE CLIMATOLOGY

ETMC-II/Doc. 5.2
(15.II.2007)

SECOND SESSION

ITEM 5.2

Original: ENGLISH

GENEVA, SWITZERLAND, 26 TO 27 MARCH 2007

PLATFORM METADATA CATALOGUE

Old versions of the WMO Ship Catalogue (WMO-No. 47)

(Submitted by Dr Elizabeth Kent¹, Mr David Berry¹, and Mr Scott Woodruff²)

*((1.) National Oceanography Centre, Southampton, United Kingdom, and
(2.) NOAA Earth System Research Laboratory, Boulder, Colorado, USA)*

Summary and purpose of document

This document provides information on the imaging and digitization status of older versions (1955-) of the WMO *International List of Selected, Supplementary and Auxiliary Ships* (WMO-No. 47).

ACTION PROPOSED

The Expert Team on Marine Climatology is invited to comment on the information provided and to consider the following items, as appropriate:

- (a) Are the Pub. 47 metadata sufficiently accessible?
- (b) What are the future priorities?

Appendices: (None)

DISCUSSION

1. Background

The WMO *International List of Selected, Supplementary and Auxiliary Ships* (WMO-No. 47, hereafter Pub. 47), has been issued since 1955. It has long been recognized that the metadata on ships and their instrumentation contained within Pub. 47 are important for the quantification and correction of bias in Voluntary Observing Ships (VOS) observations. Bias adjustment is an essential step in the process of generating climate-quality datasets from the VOS.

Because Pub. 47 was designed as an operational resource, its changing contents have never been comprehensively documented by the WMO. Therefore, a journal paper was published in February 2007, which documents the variables contained within Pub. 47 since 1955 (Kent, et al., 2007) and for the period after 1973, which links those variables to the observations made by the ships archived in the International Comprehensive Ocean-Atmosphere Dataset (ICOADS, Woodruff, et al., 1998 and Worley, et al., 2005).

Given its importance for climate research, the early paper editions of Pub. 47 have been digitized as part of the National Oceanic and Atmospheric Administration (NOAA), National Climatic Data Center (NCDC), and Climate Database Modernization Program (CDMP). The results of this digitization effort are further described in Section 2.

2. Recent Progress

2.1 Digitization of Pub. 47 Editions from 1955 to 1972

The ETMC-I was informed that editions and supplements of Pub. 47 for the period of 1955 to 1972 had been located and imaged. The only edition or supplement which had not been identified during that process (a stand-alone supplement, Edition 1966 - Supplement No. 1, October 1967), has now been found and donated to the project by Massachusetts Institute of Technology. Fortunately, it was possible for the CDMP to image this document to make it available via electronic format, and to digitize its individual supplementary metadata records. However, it was not possible for the CDMP to consolidate these records into those from the 1966 Edition, to create, in effect, a new edition.

Converting the searchable pdf images of Pub. 47 into an electronic format was a complex process. Therefore, it was necessary to define the exact process by which the information would be keyed and stored. This required several iterations, and each Edition had to be checked and sent back to the CDMP for correction(s), as appropriate or required. One unforeseen problem was the poor print quality of some of the paper copies. This problem was unable to be resolved, and it was decided to flag said entries (particularly the callsigns), which could not be read due to poor print quality.

The digitization of Pub. 47 was completed by the CDMP in early 2006.

2.2 Imaging of Pub. 47 Editions from 1973 to 1998

The WMO issued Pub. 47 as a paper publication between the period of 1955-93, and in 1994 moved to additionally providing the information via electronic format. From 1999, the electronic version was issued quarterly (but was often available only with a significant delay). Digitization of the Editions from 1973-93 was not deemed to be essential, as the word processing files used by the WMO to produce the paper editions had become available. However, it was thought useful to image the Editions for the period of 1973-98 to help with the interpretation of the metadata.

2.3 Availability of Imaged and Digitized Pub. 47 Editions

The imaged Pub. 47 Editions for the period of 1955-72 are accessible by using the Web Search Store Retrieve Display (WSSRD) software, which is available from the CDMP, located at the following web address: www.ncdc.noaa.gov/oa/climate/cdmp/wssrd.html. The WSSRD site contains all of the information imaged by the CDMP, of which the Pub. 47 metadata are only a small part. To access and use the WSSRD site, it is necessary to have an account. WSSRD accounts are available only to U.S. Government employees and their contractors, educational institutions doing environmental research, and other researchers associated with NOAA projects.

However, the imaged Pub. 47 Editions are also becoming openly available, along with the electronic metadata issued by the WMO, from the ICOADS website (icoads.noaa.gov/metadata/wmo47/).

2.4 Database of Pub. 47 Metadata

The digitization of Pub. 47, as described in Section 2.1, is only part of the process of making the metadata available. Due to the operational nature of Pub. 47, typographical errors and duplicate entries can exist, sometimes with differing metadata appearing in an edition for the same ship where it has been recruited by more than one country. Additionally, as new fields are added, it may be possible to copy the new fields to earlier editions where the other metadata matches.

A new database of Pub. 47 entries have been created and are available upon request for the period of 1955-2005. The database contains the metadata from the newly digitized files, and from the existing electronic files, with entries corrected for typographical errors, duplicate entries merged and checked for consistency and new fields extended back in time, where possible. Details on the new database can be found at the following web address: www.noc.soton.ac.uk/ooc/PROJECTS/META/index.php. Updates of the database to the end of 2006 have not yet been possible, as the metadata for the last six months of 2006 have not yet been issued by the WMO (www.wmo.ch/web/www/ois/pub47/pub47-home.htm).

2.5 ICOADS IMMA Pub. 47 Metadata Attachment

To facilitate the use of the metadata contained in Pub. 47, a subset of the metadata has been merged with the VOS observations contained in the ICOADS for the period of 1973-2005, enhancing the latest ICOADS Releases. The merged metadata plus observational records are available in IMMA format from the ICOADS website (icoads.noaa.gov). An overview of the merging process is available in Kent et al. (2007), and detailed documentation is planned for addition to the ICOADS website. Note, that the metadata attachment, as of March 2007, does not include the enhancements described in Section 2.4. It is expected that the metadata attachment for 2006 will be available by approximately May 2007.

Extension of the matching observations to metadata in the earlier periods, particularly prior to 1966, when the first telecommunicated appear in the ICOADS, is problematic due to increasingly frequent gaps in the callsign information available in the ICOADS. Even post-1966, callsign information is sometimes limited, because Global Telecommunications System (GTS) data containing callsigns could be replaced by delayed-mode logbook (e.g., IMMT) data lacking callsign information (e.g., containing national ship or hull numbers instead). It might be possible to either recover additional callsign information through comparisons of the GTS and logbook data, or to match to observations using the ship name information available in Pub. 47, and sometimes available in supplementary ICOADS data (i.e., stored within an attached original format data string, possibly of variable length). Either option would be time consuming, but potentially very valuable in terms of generating climate quality datasets for the period from 1955 to the 1970s.

3. Future Work

Remaining tasks include:

- Populating the ICOADS website with additional metadata and documentation (**action: Mr Scott Woodruff and Lubker by September 2007**);
- Incorporation into metadatabase (including merger with the 1966 Edition metadata) of the October 1967 supplement (**action: Mr David Berry and Dr Elizabeth Kent to be decided**);
- Population of the ICOADS IMMA metadata attachment for 2006 (**action: Mr David Berry, Dr Elizabeth Kent, Mr Scott Woodruff and Lubker by May 2007**);
- Update of the ICOADS IMMA metadata attachment to take advantage of improved Pub. 47 database for the period of 1955-2006 (**action: Mr David Berry, Dr Elizabeth Kent, Mr Scott Woodruff, and Lubker, tentatively by September 2007**);
- Investigation of feasibility of improving metadata matching rates in the ICOADS, using either the GTS/logbook comparisons or ship name information (**action: Mr David Berry, Dr Elizabeth Kent, Mr Scott Woodruff, and Lubker by March 2008**).

References

Kent EC, Woodruff SD, Berry DI 2007. WMO Publication No. 47 Metadata and an Assessment of Voluntary Observing Ships Observation Heights in ICOADS. *Journal of Atmospheric and Oceanic Technology* **24**(2): 214–234.

Woodruff SD, Diaz HF, Elms JD, Worley SJ. 1998. COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Physics and Chemistry of the Earth* **23**: 517-527.

Worley SJ, Woodruff SD, Reynolds RW, Lubker SJ, Lott N. 2005. ICOADS Release 2.1 data and products. *International Journal of Climatology* **25**: 823-842. DOI: 10.1002/joc.1166.

Appendices: None