ICOADS Transition Meeting Report

Monday 30 Jan. –Friday 3 Feb. 2012
NOAA Earth System Research Laboratory (ESRL)
Physical Science Division (PSD)

Update Information (16 April 2012)

Following is a report from a meeting held at the Physical Science Division (PSD) of NOAA's Earth System Research Laboratory (ESRL), to discuss the future of the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). Since the report was issued, we can report some short-term restoration of NOAA Climate Program Office (CPO) FY2012 funding for the core ICOADS work, and also for the IVAD project. This report outlines a number of specific actions and deliverables planned over the next several months to transition some activities and project infrastructure from ESRL to NCDC—and/or possibly to other new international project partners. Plans for how ICOADS development and delayed-mode processing will be managed and resourced for the longer term are still under development. Our Public Notice (available from: http://icoads.noaa.gov/) will be updated when more details about the future for ICOADS become known, or please feel free to e-mail Scott.D.Woodruff@noaa.gov or Jay.Lawrimore@noaa.gov for more detailed information. *Note: A few minor updates and redactions have been made in the following report.*

ICOADS Transition Meeting Report

Monday 30 Jan. –Friday 3 Feb. 2012
NOAA Earth System Research Laboratory (ESRL)
Physical Science Division (PSD)

Executive Summary

This is the report from a meeting held at the Physical Science Division (PSD) of NOAA's Earth System Research Laboratory (ESRL), to discuss the future of the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Woodruff et al. 2011), in light of severe budget cuts in FY2012 at ESRL/PSD plus shifts in research priorities. The purpose of the meeting was to begin exploring different options available nationally and internationally for transitioning the Research and Development (R&D) and project leadership activities for ICOADS that had, until 1 February 2012, been carried out at ESRL/PSD. Participants (see Appendix A) included PSD, NCDC, and NCAR managers, and core ICOADS project participants for all three longstanding partner organizations, i.e. NOAA—its ESRL/PSD and National Climatic Data Center (NCDC)—and the US National Center for Atmospheric Research (NCAR). Two additional scientists participated from NCDC, to help expeditiously develop plans to transition as much as possible of the very extensive marine processing capabilities and expertise available until now at ESRL/PSD.

Among the overall results of the meeting, it was agreed that, owing to known resource constraints and continuing major NOAA funding uncertainties, two contingency plans for the proposed transition activities would be developed: (i) for February-May 2012 only (**Phase A**), for which funding has been located to support the salaries of the two current ESRL/PSD project participants (Sandy Lubker and Scott Woodruff); and (ii) for a proposed additional period extending at least for June-September 2012 and ideally ongoing (**Phase B**), for which every effort should be made for NOAA management and partners to locate stop-gap funding for a smooth transition from ESRL to NCDC and/or another suitable institute, which will avoid possible critical losses of processing capabilities, software, documentation and metadata, as well as long-term institutional knowledge held by the participants currently working at ESRL/PSD.

Contents

Executive Summary
1. Status Summary: Interagency and Project-External Discussion
2. Phase A (Feb.-May): Plans for the near-real-time (NRT) transition and improvement
3. Phase B (Jun.-Sep.-ongoing): Maintaining the ability to retrospectively improve ICOADS
4. Milestones (for Phase A; to be decided for Phase B)

Appendix A: Agenda and Participants: ICOADS Transition Meeting
Appendix B: Public Notice: Termination of ICOADS Development Due to NOAA Budget Cuts
Appendix C: Budget information for ICOADS Value-Added Database (IVAD) Project
Appendix D: Research to Operations (R2O) Augmented ICOADS Data Flow
Appendix E: Additional Non-Public Webpages Hosted at ESRL/PSD
Appendix F: Tide Gauge (TG) Data Issues

1. Status Summary: Interagency and Project-External Discussion (Mon. 30 Jan. 2012)

(1) The draft Public Notice describing the ICOADS project status was accepted, and it was agreed to post it as soon as possible on relevant websites of the three current US project partners (see **Appendix B**):

   a. Janet Intrieri suggested the possibility of also publishing it more broadly, e.g. through a NOAA Press Release. However, Bill Neff's recommendation was to reconsider that possibility at a later date, pending the report from this meeting, and also in view of near-term NOAA discussions that might have a bearing on a highly visible NOAA Press Release.

   b. Note: After this decision was taken, the Public Notice (as provided in **Appendix B**) was posted 30 Jan. on relevant PSD, NCDC, and NCAR webpages.

   c. Additionally, the ICOADS partners will contact organizations with websites that have products that depend on ICOADS (e.g. ERSST) to encourage them to consider also posting the notice.

(2) Transition funding to support Sandy Lubker and Scott Woodruff (through May 2012): Some NOAA Climate Program Office (CPO) FY2012 funding for this purpose was located by NCDC, and additional funding appears to be available from the CPO-funded ICOADS Value-Added Database (IVAD) (http://icoads.noaa.gov/ivad/) project. Proposed utilization of the second year IVAD funding has been verbally agreed to by the grantees, and now the repurposing of those funds needs approval from the CPO management. Thus advance discussion between Scott Woodruff (and possibly Dave Easterling) and CPO (Bill Murray/Chris Miller) will be needed, as well as possibly a revised FY2013 workplan at a later date.

(3) Steve Worley will investigate any possibility for additional funding through the visiting scientist program at NCAR. In addition to the Cooperative Institute for Research in Environmental Sciences (CIRES) in Boulder, Dave Easterling mentioned the Cooperative Institute for Climate and Satellites (CICS-NC; Asheville) as another possible institutional resource and potential future home for a restructured ICOADS partnership.

(4) Prospects for expanded international (e.g. UK Met Office Hadley Centre) or national cooperation, e.g. with Florida State University (FSU), were noted, including the potential for assuming the project leadership from ESRL/PSD. This report will help shape the foundation for further discussions in this area.

(5) NCDC clarified their immediate near-term role, which will be limited to transitioning and improving the near-real-time (NRT) ICOADS data processing that exclusively accounts for

the modern-day time series extension with Global Telecommunication System (GTS) data. However, during the course of this meeting the NCDC visitors sought to learn the important aspects of delayed-mode processing. This significant ICOADS function involves adding newly available retrospective data sources and full data reprocessing of the multi-century historical period of record. This activity does not have any short-term or long-term future plan at this time and preparation for the next generation ICOADS Release has been stopped.

(6) NCDC agreed to back up critical ICOADS software, documentation and data holdings. This will include looking at long-term holdings at NCAR and ESRL, and preparing inventories so that the appropriate important data files are transferred to and permanently archived at NCDC.

(7) Bill Neff estimated that Sandy Lubker and Scott Woodruff would be covered through the end of May 2012 based on the above-identified NOAA/CPO funding. Sandy is provisionally to be reassigned to PSD0 (IT Web&Data) to work on NOAA AR5/CMIP5 data management. NCDC will consider whether at a later date (probably June) funding could be located by NCDC to provide coverage through FY2012 (Sept. 2012) possibly through temporary reassignment(s) to work for NCDC in Boulder, like the NOAA Paleoclimatology branch of NCDC housed here in the David Skaggs Research Center (DSRC) facility together with ESRL, or possibly at CIRES.

(8) There was also a discussion about a future home for ICOADS at NGDC or NODC. It was felt that NCDC was a much better fit, among the NOAA/NESDIS data centers, because of their long-term involvement and experience with ICOADS and the fact that the marine surface data includes many atmospheric variables similar to land surface stations.

(9) Bill Neff suggested two contingency plans be developed, (i) assuming funding for PSD personnel (ref. item 7) is available through May; and (ii) through Sept. 2012. Janet Intrieri will provide Dave Easterling the funding amounts, including overheads, required to cover Sandy and Scott through Sept. 2012.

(10) Bill Neff agreed that legacy file holdings associated with ICOADS would be preserved at PSD for at least one year, since little or no staff time is required to do so.

(11) Regarding continuing ICOADS netCDF production at PSD, Bill Neff agreed on:

a. Continued access to current netCDF Release 2.5 data via PSD.

b. Continued production by PSD of the NRT monthly summaries in netCDF format.

(12) The structure of the ICOADS website (as currently hosted by PSD) was discussed, and it did not appear difficult to transition the icoads.noaa.gov domain by reassignment within

NOAA, e.g. probably to NCDC. It was noted that NCAR has been providing mirroring[1] of the icoads.noaa.gov domain (except netCDF files) for several years, and this could be continued. However, the website domain currently also includes extensive connections with international activities, e.g. JCOMM workshops, for which the future disposition will need to be discussed in more detail. There are also some ICOADS webpages at a more preliminary stage of development available only on a test webserver at PSD, which are not yet publicly available and not mirrored to NCAR, but should be considered as part of the transition effort (see **Appendix E**).

(13) It was agreed that a written report would be issued from the meeting.

2. Phase A (Feb.-May): Plans for the near-real-time (NRT) transition and improvement

Introduction: Compared with ICOADS delayed-mode processing, the NRT, or "preliminary" processing (i.e. since it is preliminary to a complete Release) is a less complex processing, currently based on just one GTS source, i.e. NOAA National Centers for Environmental Prediction (NCEP) data. This is less of an R&D effort than the delayed-mode processing (see **Phase B**)—but equally important to support a wide variety of additional climate applications—and thus more manageable as a routine operation. However, this monthly processing does have a requirement to be completed in 3-5 days from the beginning of each new month to support timely climate monitoring and production of derived products.

This meeting proposes the following actions for the transition and improvements in the NRT ICOADS processing:

i. For the IMMA format (see Table 1), NCDC to discontinue all use of the NCDC-variant format ($IMMA_v$), and adopt the standard format ($IMMA_s$) currently used for Release 2.5 (R2.5) data and the ICOADS NCEP-based NRT data.

ii. In providing ICOADS data to the public, NCDC to continue serving only the "enhanced" (ENH) subset of IMMA data. ENH means (see also Table 1) that a precisely defined set of platform type (PT) and quality control (QC) flags, stored in the IMMA report, are applied to the archive, such that reports from selected (e.g. near-coastal) platform types, and individual field values flagged as suspect, are removed from the IMMA collection. This proposed policy could be revisited e.g. if

1 The mirroring activity consists of a full backup of the relevant files executed currently every week, such that the alternative public URL http://icoads.ucar.edu could be activated if needed. At critical times the NCAR updates are run daily.

NCDC's data access services were more adaptive and allowed user selection of the ENH product similar to what NCAR provides. There is also a defined "standard" (STD) trimming that is not offered through NCDC, but can be obtained from NCAR (see Table 1).

iii. To continue to have no products offered from any project partner labeled "ICOADS" unless agreed among the partners to be official "ICOADS." Thus e.g. NCDC "global marine data," that are not yet officially part of ICOADS (e.g. relevant near-real-time and delayed-mode data now provided regularly to NCAR from NCDC, in advance of delayed-mode blending) should not be labeled "ICOADS."

Table 1. Explanation of differences in the IMMA format (ICOADS 2010) between its standard ($IMMA_s$) and NCDC-variant ($IMMA_v$) forms, and of the "enhanced" (ENH) and "standard" (STD) IMMA data subsets, versus the complete, or "untrimmed" (UNT), IMMA data. The ENH and STD definitions also apply to the production of ICOADS monthly summary statistics (basic gridded) products (ref. http://icoads.noaa.gov/e-doc/R2.5-stat_doc.pdf).

Abbreviation	*Explanation (see also Supplement E in ICOADS 2010)*
$IMMA_s$	Standard IMMA format, with the structure: C0 (core) + C1 + C2 + C3 + C4 + C6, such that the IMMA uniform data representation standard is followed, and space is saved through omission of empty attachments (i.e. information not relevant, or not available, for a given dataset).
$IMMA_v$	NCDC-variant of the IMMA format, with the structure: C0 (core) + C1 + C2 + C3 + C6, i.e. only selected attachments (e.g. omitting ship metadata) and additionally such that: (i) Date and time are represented with leading zeros (e.g. *YR, MO, DY,* and *HR:* "200707010000"), in conflict with the otherwise uniform numeric format model of IMMA, which has no leading zeros (e.g. "2007 7 1 0"); (ii) Longitude is expressed according to the NCDC convention (−179.99° to 180.00°E) as opposed to the $ICOADS_s$ convention (0.00° to 359.99°E); and (iii) a national ID (*NID*) field is set differently for some data distributed by NCDC.

Abbreviation	*Explanation (see also* http://icoads.noaa.gov/e-doc/stat_trim*)*
ENH	Enhanced subset: The enhanced data were derived using 4.5σ trimming (QC) limits, so as to accommodate more extreme climate events, and using a broad collection of marine observations including ships, buoys, and near-surface

oceanographic profile temperatures. These are the most widely used data, and what are presently distributed by NCDC, i.e. $IMMA_v$-EHN. However in the future it is proposed that NCDC will distribute $IMMA_s$-ENH data, as are also distributed by NCAR via their subsetting interface.

STD Standard subset: The standard data were derived using 3.5σ trimming limits and were limited as nearly as possible to observations taken from ships (where identifiable)—which data are distributed by NCAR via their subsetting interface.

UNT Untrimmed IMMA data. These are the complete data as distributed by NCAR ($IMMA_s$-UNT) without any report eliminations (for purposes of platform type subsetting) and without any field deletions (through the application of trimming flags to eliminate suspect/erroneous data)

With that groundwork in place, we suggest that NCDC management consider four different scenarios for the near-term transition of ICOADS NRT updates, which would have different resource requirements and commitments:

I. Minimal: NCDC to simply regularly ingest the IMMA NRT NCEP data currently prepared monthly by NCAR, and serve the ENH subset of those data to the public in the standard IMMA form (i.e. $IMMA_s$-ENH).

II. Intermediate: NCDC to run in parallel the processing of the NCEP data, to ensure that the processing is running correctly and identically at both NCAR and NCDC. This would help facilitate a transfer of the processing to NCDC from NCAR in the future if deemed desirable, and meanwhile provide redundancy. This also serves the data to the public as outlined in option I.

III. Advanced and recommended: This adds translation of the NCDC GTS data into the IMMA format, and then blends it with the NCEP GTS data as in option I. or II. This will involve the development of a new duplicate elimination (dupelim) processing step that is a scaled down version of dupelim described in Phase B below (see also: http://icoads.noaa.gov/dupelim.html). This option maximizes the availability of unique observations from both sources and creates a superior product compared to what is now available from NCDC or NCEP.

IV. Semi-delayed-mode (ref. **Appendix D**): Longer-range possibility, with a close relationship with **Phase B** processing: Include additional NRT and delayed-mode sources (e.g. US and international keyed VOS, and SEAS data), and perform frequent

updates for fairly recent historical periods as new data become available (similar to an operational product like GHCN).

In the event that the datastream(s) from NCEP and NCDC change substantially (e.g. WMO's previously mandated migration to BUFR in 2012), NCAR and NCDC will jointly work to resolve the effect of the changes and restore ICOADS NRT production to operations.

The major actions needed to realize the goals of the recommended NRT scenario are:

1. Assure that NCDC climate data on-line access (CDO) to R2.5 ENH data is operating properly.

2. NCDC to establish a routine process to get monthly ICOADS NRT (NCEP GTS-based) updates from NCAR. Also obtain from NCAR all past NRT updates 2008 forward. Then NCDC will be serving uniformly, in parallel with NCAR, both R2.5 and NRT data—i.e. both official components of ICOADS. Several actions will be required including:

    a. Resurrect the R2.5 regular (non-variant format) IMMA data copy from NCDC archives, and convert it to the ENH subset in the standard IMMA format (thus yielding $IMMA_s$-ENH).

    b. Similarly, the ICOADS NRT data will be obtained monthly from NCAR, archived, and filtered to create the ENH subset (thus yielding $IMMA_s$-ENH), which will be made available by NCDC to the public.

3. NCDC to also work with NCAR to obtain copies of: (i) currently operating ICOADS NRT processing software, and (ii) scripts. Also NCDC will study the data flow for the ICOADS NRT (NCEP GTS-based) data.

4. PSD will investigate providing NCDC with a stripped down version of a delayed-mode duplicate elimination (dupelim) program (with an overview provided at http://icoads.noaa.gov/dupelim.html, and with the very complex dupelim and associated rules for one particular Release and time period shown for example here: http://icoads.noaa.gov/e-doc/other/dupelim_1980), which should be more suitable for adaptation for the proposed NRT blending of the NCEP and NCDC GTS data.

5. NCDC should ensure availability of the necessary computer resources, which however are anticipated to be more manageable for the NRT processing.

6. At NCDC, the GTS decoder component of the general marine processing program (as used for processing all marine observations) will be replaced by the PSD-developed GTS decoder. This will output standard $IMMA_s$ and resolve some known decoding problems

in the NCDC software. Future merging of NCEP and NCDC GTS data streams will then be more directly achievable.

7. Other datastreams flowing operationally from NCDC to NCAR will also gradually be transitioned away from the variant IMMA format (i.e. to $IMMA_s$ from $IMMA_v$), as resources permit. The ultimate goal is to abandon $IMMA_v$ altogether at NCDC.

8. The most recent comparison of NCDC and NCEP GTS data showed data differences between the two sources http://icoads.noaa.gov/deck99x.html, which need to be resolved through further validation checking. But this work, and also broader GTS intercomparison work completed previously by Dave Berry and Liz Kent (UK NOC) for ETMC (http://icoads.noaa.gov/etmc/ETMC3-GTS_Comparison_revised2.pdf) showed clearly the value of merging multiple GTS streams to maximize data availability and data quality. Some specific benefits of the proposed merging and blending of the NCDC and NCEP GTS data include:

    a. Capturing ship ID information missing since Dec. 2007 in the NCEP GTS data, but more frequently available in the NCDC GTS data, thus allowing the critical Voluntary Observing Ship (VOS) metadata (WMO Pub. 47) to be associated with the individual marine reports (ref. **Appendix D** and Berry et al. 2009).

    b. NCDC GTS appears to have more Coastal-Marine Automated Network (C-MAN) data (possibly owing to higher temporal resolution data, or different stations).

    c. The representation of Tide Gauge (TG) stations in both NCDC and NCEP GTS differs, with only NCEP offering both C-MAN-format and CREX-format stations (NCDC has only the C-MAN-format type) (see **Appendix F**).

    d. For ship data, for example, studies have indicated (as mentioned above) that it is likely that if blended together both NCEP and NCDC GTS would provide some additional unique observations.

    e. However, it was emphasized that the important role of NCEP's GTS "dup-merge" program (as described in Appendix A of http://icoads.noaa.gov/rt.html) needed to be kept in mind in interpreting any of the NCDC-NCEP intercomparisons. Essentially the NCEP dup-merge program adds value to the GTS data by compositing and rescuing some records corrupted in transmission. This can result in fewer total records, but with a higher quality overall. This process is not run at NCDC and therefore would appear in a comparison test. It was suggested it be investigated whether NCEP's dup-merge code could be obtained for possible reuse in this project at NCDC.

9. As part of this improved NRT processing, the important possibility of a reprocessing of the GTS data for ICOADS should be considered as resources permit, using both NCDC and NCEP GTS data, at least back to Dec. 2007 in order to retrospectively resolve the ship call sign masking problem that began then.

10. The general issue of the C-MAN location list needs to be resolved, so that the locations in the NCDC list are improved in resolution. This has been the subject of some discussion with NDBC (Dick Crout) as to the desirability of creating a definitive list for use by the user community as a whole. The best result for ICOADS would be the creation of a station history that covers as much time retrospectively as possible.

11. As resources permit, it is suggested that the processing and proposed blending of two independent GTS datastreams (i.e. NCDC and NCEP) also offers the opportunity for enhanced data flow monitoring, seeking to detect problems in either GTS stream (see e.g. Figure A1 http://icoads.noaa.gov/deck99x.html).

12. From the activities above we will have created two identically formatted near-real-time GTS streams, one from NCDC and one from NCEP. The next step would be to do a careful intercomparison of both datastreams, identifying the data characteristics that are in common and different (building on previous studies, e.g. item 11). Understanding of these characteristics will lead to well-formulated dupelim rules, and ultimately to a single blended GTS datastream to become the improved ICOADS NRT.

13. In designing the data characteristics study, collaboration will be planned between PSD and NCDC so as to leverage the extensive past experience from PSD.

14. The project partners intend to run this operationally and prove it is stable and reliable for a considerable period (e.g. several months) before it replaces the existing NCEP-based preliminary ICOADS NRT data.

15. Access services from NCDC from this effort will first focus on monthly files (ftp access only), later to be followed by access through NCDC's CDO, based on the standard (non-variant) $IMMA_s$ format. NCAR will continue to serve the official ICOADS archive as it has in the past. Additionally NCDC plans to present the directions of this project to CDO developers and data access managers at NCDC in the near future, so that transition and changes within the CDO can be most effectively implemented.

16. Currently and in line with the above plans NCDC would serve only the enhanced subset of ICOADS observations (i.e. $IMMA_s$-ENH), but this should be revisited for discussion among project partners. We note that there are two reasons this decision has been taken in the past: (i) so NCDC will offer unique data access services not offered by other

project partner(s), and (ii) so that users will not be served "untrimmed" data (i.e. raw observations without QC flags applied, typically used only by highly experienced ICOADS users).

17. As a secondary priority (under **Plan B**) the possibility of expanding ICOADS to include meteorological measurements from TG stations (both CREX and C-MAN-formatted) should be explored, as well as the possibility (linked with very extensive IMMA changes proposed under IVAD) to extend IMMA to include the TG water level data (see **Appendix F**).

18. Following successful validation and production of a merged NRT product, NCDC and NCAR will consider the possibility of eventually moving full responsibility for the NCEP GTS ingest and processing to NCDC (only after a long period of overlap however).

19. Milestones as presented in section 4 are intended to transition the ICOADS NRT processing as appropriate to NCDC, and if possible produce an improved official NRT product, by the end of May 2012.

3. Phase B (Jun.-Sep.-ongoing): Maintaining the ability to retrospectively improve ICOADS

Introduction: Retrospective processing for ICOADS is an extremely complex process that brings together order 100 different international marine data sources into a uniform homogeneously treated collection. Each data source involves unique issues: such as different observing systems and practices over 300 years; different national and international coding schemes, and many unique formats requiring carefully validated translations; matching different "duplicate" records, and selecting the most valuable and complete; etc. This kind of effort is more of a research and development effort, so that the activity is not well suited for operations. Further, the updates for the historical period cannot be executed in near-real-time, because they usually require a few iterations—separated by periods of study and in-depth investigation on impacts from new data sources.

Over the past three decades ICOADS has become the worldwide respected marine surface dataset. ICOADS is the foundation for most internationally developed global ocean surface products, analyses, and reanalyses. As such the international community has relied on the data management/stewardship leadership and data synthesis expertise from NOAA/ESRL project members, and on the ICOADS archive as the home for data rescue and improvement internationally. Over time ICOADS has had about 10 major Releases and each one has resulted in ever improving climate research data sets. The current ICOADS expertise can never be replaced, and therefore this transition plan recommends specific actions and resources to preserve this international data treasure and offer a potential successful path into the future.

Activities related to delayed-mode processing and data stewardship, and that result in a new ICOADS Release:

1. Gathering of new or improved data sources.
2. Format homogenization and validation, and study of individual data source characteristics. Includes detailed documentation for each format translation.
3. From the studies of data characteristics, and also knowledge gleaned from user feedback on the previous Release, rules are compiled to carefully blend together the large number of input data sources, correct known errors, and remove unwanted duplication between data sources.
4. Software development, testing, and validation of the rules specifically determined through the above steps for a given new Release.
5. Iterative processing using that software, followed by in-depth evaluation of the output.

6. Adapt and refine the rules and software to adjust for detected unrealistic data characteristics.

7. Publication of data and user documentation, including website updates, user data access GUI upgrades, and community notification.

8. Develop more detailed documentation of all aspects of the new Release, e.g. data characteristics, errors discovered, and solutions, if available, to correct those errors—together with a citable journal publication so the Release can be conveniently understood and referenced.

9. Publication of the Release—data and documentation, including website updates, and user outreach.

International coordination and data management functions:

1. JCOMM Expert Team and Task Team activities, including the JCOMM Expert Team on Marine Climatology (ETMC) (http://icoads.noaa.gov/etmc/) (Current Chair: S. Woodruff)

2. JCOMM-organized marine workshops, MARCDAT and CLIMAR

3. The new "CLIMAR" umbrella international initiative planned to manage future such marine climatology workshops (http://www.marineclimatology.net/web/) (Current Chair: S. Woodruff)

Recommendations looking to the future:

1. It is critical to carry out one new Release with full or partial involvement of the current ESRL experts. This will afford a minimal education for a new retrospective data processing team. Release 2.6.0 was originally planned for completion by October 2012, and had essential components for the success of the new IVAD project. Importantly, IVAD was planned to create an adjusted ICOADS that accounted for observing system changes and inhomogeneities introduced by simultaneous operating observing systems, e.g. ship and drifting buoys. At this time all progress toward a new ICOADS Release and IVAD have been stopped.

2. Within the transition meeting agenda (see **Appendix A**), under the Project-internal discussion portion, a large number of detailed logistical considerations are provided, concerning ICOADS past input data files, software, documentation and metadata holdings, and key project webpages (either available to the public, or not yet generally available). These should be referenced in the design of further detailed plans for **Phase B**.

3. The model for the ICOADS development has been a huge success, enduring for decades and resulting in the world's open reference dataset for the marine surface. Future efforts should critically review the ICOADS history and approach, understand how the program has made these data so valuable, and consider how to sustain this development within the internationally responsible community. Sudden and very real budget problems will always have impact, but some resiliency for a data treasure such as ICOADS should be better planned.

4. Milestones

Phase A (Feb-May 2012)

1. Review the draft meeting report and discuss with NCDC management the near-real-time (NRT) transition/improvement, and the specific scenario to be chosen by NCDC **(10 Feb.)**.

2. ESRL to provide an image of the icoads.noaa.gov domain to NCDC for technical review **(completed 2 Feb.)**.

3. NCDC to obtain the backlog of NCEP preliminary NRT data (2008-Jan 2012) **(completed 6 Feb.)**.

4. NCDC to initiate submission agreement process to archive the NCEP NRT data **(17 Feb.)**.

5. ESRL, after receiving from NCDC the program currently used, to modify the program that produces the enhanced subset, after revalidating the output against NCAR output format, from producing $IMMA_v$ to produce the standard $IMMA_s$ format, and do some checks **(completed 7 Feb.)**.

6. NCDC to convert R2.5 plus NRT update backlog data into enhanced subset, and publish them to a new NCDC ftp site for public access **(24 Feb.)**. Related, NCDC will discontinue the ftp access to the variant-formatted ICOADS R2.5 that is currently available at NCDC.

7. NCDC to begin operationally receiving ongoing updates via "push" mode ftp from NCAR of the NCEP NRT data **(Mar.)**. Note: review and establish checksum security (both NCDC and NCAR). This involves writing script code to do the following:

    a. Detect arrival of new file **(2-5 Mar.).**

    b. Submit file received from NCAR to NCDC permanent archive (upon establishment of submission agreement).

    c. Process file to convert to enhanced subset version (using software from step (3)).

d. Publish enhanced subset file (from c) in designated ftp area (ref. step 4)

e. Submit enhanced subset file to archive (again pending submission agreement).

f. Submit file created in (e.) to NCAR for additional permanent archival.

8. NCDC to discontinue usage of the $IMMA_v$ format, beginning with the NRT NCEP data and R2.5 **(ASAP)**.

Assuming the recommended Scenario III **(Mar.-May 2012)**:

9. NCAR software provision of the existing NRT NCEP codes to NCDC, for study at NCDC.

10. Develop a process to rectify differences between NCEP and NCDC C-MAN station locations, including further consultation with NDBC (Dick Crout).

11. Complete installation at NCDC of the ESRL-developed GTS decoder, to produce $IMMA_s$ formatted data.

12. ESRL to provide suitable scaled-down dupelim program, as the starting point for development by NCDC in collaboration with ESRL of a final operational NRT dupelim program to merge both the NCEP and NCDC GTS datastreams.

13. NCDC to develop programs and strategy to reevaluate the data characteristics of NCEP and NCDC GTS data, as well as the characteristics of the merged GTS product, in collaboration with ESRL. This is anticipated to be an iterative process, to make improvements in the dupelim program from step 12.

14. Ideally, transition by May 2012 to the blended product as the new and improved ICOADS NRT product.

15. In that event, submission agreements required for NCDC archival, etc.

16. In that event also, NCAR to serve these improved blended data instead of existing NCEP NRT data.

17. Potentially move the NCEP processing from NCAR to NCDC, if so desired.

Phase B (June-September 2012, and ongoing)

The transition meeting was unable to progress in any detail on the development of detailed milestones for **Phase B**, which actions in any case may be strongly dependent on the path chosen for the future for replacing the leadership role that has been held by ESRL/PSD, as well as taking over the very substantial delayed-mode and historical data processing role. It was

noted however that some task elements and considerations as listed in section 2 as under **Phase A** (e.g. possible utilization for ICOADS of surface meteorological data regularly reported from TG stations, or reprocessing of the NCDC GTS data back to at least December 2007 seeking to better resolve the callsign masking problems that exist in the NCEP GTS data starting around then) would probably also realistically fall under **Phase B** in terms of timing.

References

Berry, D.I., E.C. Kent, and S.D. Woodruff, 2011: Blending ICOADS Release 2.5 and WMO Publication 47, 1966-2007 [http://icoads.noaa.gov/e-doc/imma/WMO47IMMA_1966_2007-R2.5.pdf].

ICOADS, 2010: Archival of Data Other than in IMMT Format: The International Maritime Meteorological Archive (IMMA) Format. [note: full version including extensive background available at http://icoads.noaa.gov/e-doc/imma/R2.5-imma.pdf or abridged version limited to format details at http://icoads.noaa.gov/e-doc/imma/R2.5-imma_short.pdf].

Woodruff, S.D., S.J. Worley, S.J. Lubker, Z. Ji, J.E. Freeman, D.I. Berry, P. Brohan, E.C. Kent, R.W. Reynolds, S.R. Smith, and C. Wilkinson, 2011: ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* (CLIMAR-III Special Issue), **31**, 951-967 (doi:10.1002/joc.2103).

Appendix A

Agenda and Participants: ICOADS Transition Meeting

Monday 30 Jan. –Friday 3 Feb. 2012 (note: ending by 11am Friday)

NOAA Earth System Research Laboratory (ESRL)

Physical Science Division (PSD)

Room 3B207

Participants **(note: largely limited to NCDC visitors and core project participants after Monday 30th)**

NOAA/ESRL/PSD:	**NOAA/NCDC:**	**NCAR:**
Don Hooper ("Hoop"; PSD Web&Data alternate)	Dave Easterling (E/CC1 Chief; *telecon. only*)	Steve Worley (Manager, CISL Research Data Archive)
Janet Intrieri	Eric Freeman	
Sandy Lubker	Boyin Huang	**FSU:**
Bill Neff (PSD Director)	Jay Lawrimore (E/CC11 Chief; *telecon. only*)	Shawn Smith (Chair SAMOS and Data Center Dir.; *telecon. only*)
Cathy Smith (PSD Web&Data)	Dave Wuertz	
Jim Wilczak		**NOAA/NMSP:**
Scott Woodruff		Catherine Marzin (NOAA National Marine Sanctuaries Program; *telecon. only*)

Meeting contacts:

Scott Woodruff 325 Broadway; Boulder, CO 80305 USA Phone: 303-497-6747 (fax: -6181) Scott.D.Woodruff@noaa.gov	Administrative support contact: Madeline J. Sturgill, PSD3 Secretary Phone: 303-497-5961 (fax: -6181) Madeline.Sturgill@noaa.gov

Background (excerpted from Public Notice; see Appendix B):

ICOADS has operated for decades in the US as a partnership between NOAA—its Earth System Research Laboratory (ESRL) and the National Climatic Data Center (NCDC)—and the National Center for Atmospheric Research (NCAR), together with growing international contributions.

For budgetary reasons, stemming from pending large cuts at the NOAA Climate Program Office (CPO), ESRL Directors have determined that it is no longer feasible for its Physical Science Division (PSD) to continue supporting any further ICOADS work—effective immediately.

This immediate cessation of support by ESRL means the future of ICOADS is unclear. The program is working to transition some responsibilities to NCDC, but for the foreseeable future improvements to the historical time period are stopped. For the near-term, data access services of the existing Release 2.5 (1662-2007) will continue through NCAR and NCDC (and possibly PSD).

Monday, 30 January 2012

Interagency and Project-External Issues

8:30-10:00am M.S.T.	A. Higher-level transition and data management topics: **1.** Termination of support for ICOADS at ESRL/PSD Funding and institutional background (NCDC and PSD) Prospects for expanded international cooperation (e.g. Hadley Centre) **2.** Draft Public Notice (see **Appendix B**) regarding termination—Status of PSD internal review **3.** Prospects for residual continuing transition activities at ESRL/PSD (a) Reassignment status (Lubker, 1 Feb. to PSD0; Woodruff, pending) (b) Proposed retention for ≥1-year of ICOADS legacy file holdings (including e-mail, and /home and /data) for continuing access and transition activities **4.** Fate of the ICOADS Value-Added Database (IVAD) project (~$31.5K currently slated for PSD in FY12; ref. **Appendix C**) (http://icoads.noaa.gov/ivad/) **5.** ICOADS netCDF access/production questions: proposed treatment starting immediately of ICOADS as an ordinary "external" data set offered by PSD, as follows: (a) PSD to offer continued access to ICOADS current Release 2.5 (1662-2007)—static netCDF files (b) PSD to offer continued access to ICOADS "preliminary" real-time extensions (2008-forward)—NCAR processes these observations and summaries, but this would then involve ongoing (and timely) turnkey production by PSD of netCDF (~2-5 days after the end of each data month) **6.** Disposition of the icoads.noaa.gov website domain, and content i. icoads.noaa.gov domain could presumably be transitioned elsewhere within NOAA, e.g. to NCDC, if desired ii. icoads.noaa.gov content (but *not* netCDF files as served by PSD) has been regularly mirrored for several years by NCAR, to guard against catastrophic loss and provide back-up capability (e.g. in the event of US Gov't shutdown) **7.** Agreement on the form of a report from this meeting.
10:30-11:00	Coffee Break
11:00-12:00	A. Higher-level transition and data management topics (continued, as needed)
12:00-13:00	Lunch
13:00-15:00	B. Technical discussion related to above transition topics
15:00-15:30	Coffee Break
15:30-17:00	B. Technical discussion related to above transition topics (continued)

Tuesday, 31 January 2012

Project-internal discussion

9:00-10:00	C. Detailed transition logistics **1.** Input/output data and file issues (a) legacy data files, i.e. from previous Releases (available at PSD and NCAR, e.g. as early as utilized for COADS Release 1 c. 1985) (b) "tapelists" (old lists of file volumes used for earlier Releases) (c) current input, "intermediate" output (i.e. containing all dups), and output files from Release 2.5 (1662-2007) (d) new "Auxiliary" data inputs (in IMMA format, e.g. German Maury Collection, GODAR, deck 117, EEIC, ref. item 10; or otherwise available in original input formats) **2.** General software questions (a) the desirability of exposing legacy/current to the general public for possible free re-use and examination. (b) note: many legacy programs currently may not be sufficiently branded as to original authorship (e.g. by PSD or NCAR) (c) program languages used (Fortran mostly) (d) compilers (e) shell scripts **3.** Hardware (a) NCAR and ESRL servers presently used for ICOADS processing i. make(s) ii. operating systems iii. capacity **4.** "Preliminary" real-time processing (a) translation of NCEP GTS—future homogeneous translation of NCDC GTS (same decoder, as developed at PSD?) (b) duplicate elimination (dupelim)—not applicable currently, but a simple dupelim was planned under an enhanced operationalization proposal that had been under active discussion among project partners, to also utilize NCDC GTS data (in addition to NCEP). (c) QC (ref. http://icoads.noaa.gov/e-doc/stat_trim, Note: this documentation is out of date). (d) monthly summaries: 2° latitude x 2° longitude boxes back to 1800 (and 1°x1° boxes since 1960): standard and enhanced (also special equatorial products) processed into binary MSG format (http://icoads.noaa.gov/e-doc/msg). (e) translation of monthly summaries from MSG into netCDF format (current PSD responsibility) **5.** Delayed-mode (full Release) processing (a) translation of many data inputs (b) dupelim: very complex processing (overview provided by Worley to International Surface Temperature Initiative (ISTI) available here:

	http://icoads.noaa.gov/dupelim.html), requiring in-depth tuning and iterative test processing for each new Release, to adjust to new data input characteristics, etc. Typically different "preconditioning" and dupelim rules have been required for different time periods (e.g. http://icoads.noaa.gov/e-doc/other/dupelim_1784 versus http://icoads.noaa.gov/e-doc/other/dupelim_1980). (c) QC (as under 4) (d) monthly summaries (as under 4) (e) netCDF (as under 4) **6.** Research to Operations (R2O): Operationalizing delayed-mode and near-real-time ICOADS processing, similarly to the Global Historical Climatology Network (GHCN) (a) proposed merger (and simple dupelim) of NCEP and NCDC GTS (http://icoads.noaa.gov/deck99x.html). (b) previous roadblocks/concerns; **Appendix D** describes some ideas previously under discussion, focused primarily on improved handling of the GTS data. **7.** Data characteristics and unresolved data problems (a) R2.5 data characteristics: http://icoads.noaa.gov/r2.5.html (b) SEAS data problems: http://icoads.noaa.gov/deck874.html (c) NDBC moored buoy and C-MAN historical data archive issues (w/ NODC) (d) Numerous additional outstanding data problems/questions: http://icoads.noaa.gov/pre2.5-2.html **8.** Platform and instrumental metadata (a) WMO Pub. 47 Voluntary Observing Ship (VOS) historical metadata holdings at PSD: http://icoads.noaa.gov/metadata/wmo47/ (b) Blend of Pub. 47 metadata with ICOADS, e.g. for R2.5 this was implemented in close cooperation with UK NOC: http://icoads.noaa.gov/e-doc/imma/WMO47IMMA_1966_2007-R2.5.pdf **9.** Documentation considerations (a) e-doc (http://icoads.noaa.gov/e-doc/) Important outstanding issue: some key R2.5 documentation, as also noted above remains only partially completed, including dupelim documents for different time periods, and a long overdue update of the "stat_trim" documentation describing QC flags. (b) transpec: field-by-field format translation documentation that has been written in some cases to accompany translation programs, e.g.: http://icoads.noaa.gov/e-doc/other/transpec/maury/maury_transpec. Originally, when resources were more abundant (COADS Release 1, c. 1985) translations were implemented by two people independently, which could utilize the transpec as the "common ground" for the back-up translation, and then results could be mechanically cross-verified (and differences of interpretation, or simply programming errors, were invariably found). (c) ICOADS web info pages such as these for the real-time and delayed-mode archives, respectively (both pre-R2.5): http://icoads.noaa.gov/rt.html

http://icoads.noaa.gov/r2.html

(d) Translation webpages (using including transpec, plus other background information and examination results), http://icoads.noaa.gov/translation.html

10. Web-presence/webpage questions

(a) Webpages not yet migrated to full public availability (see also **Appendix E**):

http://icoads.noaa.gov/ivad/
http://icoads.noaa.gov/20th.html
http://icoads.noaa.gov/maury_german.html
http://icoads.noaa.gov/godar.html
http://icoads.noaa.gov/deck117.html
http://icoads.noaa.gov/noaarv.html
http://icoads.noaa.gov/dupelim.html
http://icoads.noaa.gov/deck700.html
http://icoads.noaa.gov/deck874.html
http://icoads.noaa.gov/deck99x.html
http://icoads.noaa.gov/eeic.html [Note: IMMA data provided by the Hadley Centre, but they were planning to provide a reconversion to fix some small problems]

(b) JCOMM Events/Information

CLIMAR (I-III) and MARCDAT (I-III) workshops:
Expert Team on Marine Climatology (ETMC): http://icoads.noaa.gov/etmc/

11. RECovery of Logbooks and International Marine data (RECLAIM) project: http://icoads.noaa.gov/reclaim/

i. Marine data rescue document (http://icoads.noaa.gov/reclaim/pdf/marine-data-rescue.pdf).
ii. Many historical items imaged by CDMP were planned for addition to the RECLAIM website, but time has not permitted them to be added (at least a list of these could be useful, and identification of their location).
iii. Before he was abruptly terminated by CDMP, Clive Wilkinson had (and should still have) many additional materials (e.g. archive inventories) he was planning to add to the RECLAIM website, under a planned restructuring and refurbishment. A document is available describing the website redesign plans.

12. JCOMM and other international coordination issues

(a) Future US representation on the JCOMM Expert Team on Marine Climatology (ETMC).

(b) Continuing US representation on the JCOMM Task Team on the Marine Climate Data System (MCDS).

(c) Discussion regarding the previously planned US application for ICOADS to become a WMO-IOC Centre for Marine-meteorological and Oceanographic Climate Data (CMOC). Note: Application for JCOMM-IV due 15 Feb. but the prospects are probably nil now, considering the ICOADS and NOAA resource situation.

(d) The new "Climar" overarching marine climatology workshop initiative (S. Woodruff currently Chair): http://www.marineclimatology.net/web/

Note: S. Woodruff also currently a Member (2001-) of the GCOS (AOPC/OOPC) Working Group on Surface Pressure (WG-SP) and of the OceanSITES Data

	Management Team.
10:30-11:00	Coffee Break
11:00-12:00	C. Detailed transition logistics (continued)
12:00-13:00	Lunch
13:00-15:00	C. Detailed transition logistics (continued)
15:00-15:30	Coffee Break
15:30-17:00	C. Detailed transition logistics (continued)

Wednesday-Friday, 1-3 February 2012 (note: ending by 11am Friday)

Project-internal discussion (continued as needed)

9:00-10:00	C. Detailed transition logistics (continued)
10:30-11:00	Coffee Break
11:00-12:00	C. Detailed transition logistics (continued)
12:00-13:00	Lunch
13:00-15:00	C. Detailed transition logistics (continued)
15:00-15:30	Coffee Break
15:30-17:00	C. Detailed transition logistics (continued)

Appendix B

Public Notice: Termination of ICOADS Development Due to NOAA Budget Cuts

30 January 2012

(NOTE: date of posting on ICOADS websites at ESRL/PSD, NCDC, and NCAR)

ICOADS has operated for decades in the US as a partnership between NOAA—its Earth System Research Laboratory (ESRL) and the National Climatic Data Center (NCDC)—and the National Center for Atmospheric Research (NCAR), together with growing international contributions.

For budgetary reasons, stemming from pending large cuts at the NOAA Climate Program Office (CPO), ESRL Directors have determined that it is no longer feasible for its Physical Science Division (PSD) to continue supporting any further ICOADS work—effective immediately.

This immediate cessation of support by ESRL means the future of ICOADS is unclear. The program is working to transition some responsibilities to NCDC, but for the foreseeable future improvements to the historical time period are stopped. For the near-term, data access services of the existing Release 2.5 (1662-2007) will continue through NCAR and NCDC. Currently it is anticipated that:

- At this juncture there are no plans for any new major ICOADS delayed-mode updates or further Releases.
- "Preliminary" near-real-time updates of the observations and basic gridded monthly summary products will continue to be produced and served at NCAR for as long as current data flows remain unchanged.
- The ICOADS website (http://icoads.noaa.gov/), the project central information source, as hosted at PSD cannot be updated further. A future home for the extensive existing content (including documentation, software, and metadata) is under discussion.
- This situation extends to related websites currently hosted under the icoads.noaa.gov domain, including the international marine climatology workshops MARCDAT and CLIMAR, as well as the recently funded ICOADS Value-Added Database (IVAD) Project. The IVAD project is hinged on developments at ESRL, so it is likely this will be terminated.
- Regrettably moreover, there will be no staffing support for marine work or consultation from ESRL.

This situation suddenly became known to the project membership in December 2011, and we are very sorry about this development and for such short notice. We know it will have significant impacts on the quality of a wide range of scientific research. Please address any further questions or concerns to ESRL, NCDC, and/or higher-level NOAA management.

Appendix C
Budget information for ICOADS Value-Added Database (IVAD) Project

Note: possible reduction equally across all four organizations by 25% for FY2012, as a potential response to NOAA/CPO guidance (ref. 21 Dec. 2011 e-mail from Bill Murray) that the overall budget must be reduced somehow by 25% for FY2012:

Table 3. Summary of project costs separately by institution and totals.

Funding year	ESRL/PSD	COAPS/FSU	NCAR	NCDC	Total
FY2011	$64K	$23K	$33K	$32K	$152K
FY2012	~~$42K~~	~~$66K~~	~~$23K~~	~~$21K~~	~~$152K~~
	$31.5K	$49.5K	$17.25K	$15.75K	$114K
FY2013	$41K	$69K	$21K	$21K	$152K
Total:	<$147K	<$158K	<$77K	<$74K	**<$456K**

S. Woodruff, 18 January 2012

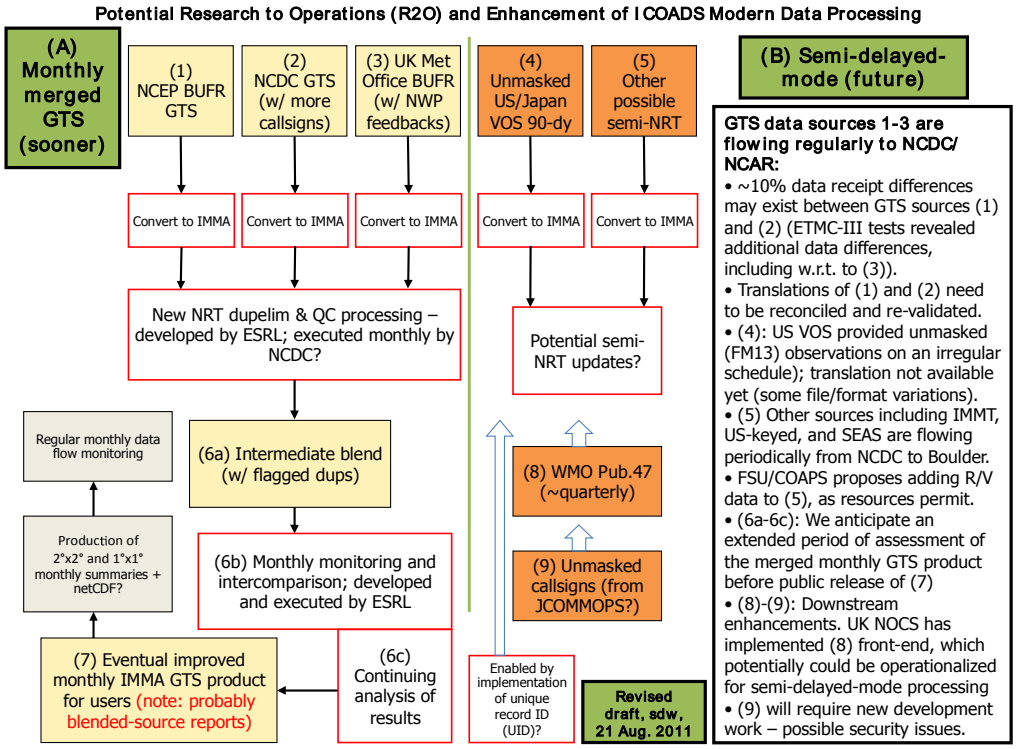
Appendix D

Research to Operations (R2O) Augmented ICOADS Data Flow

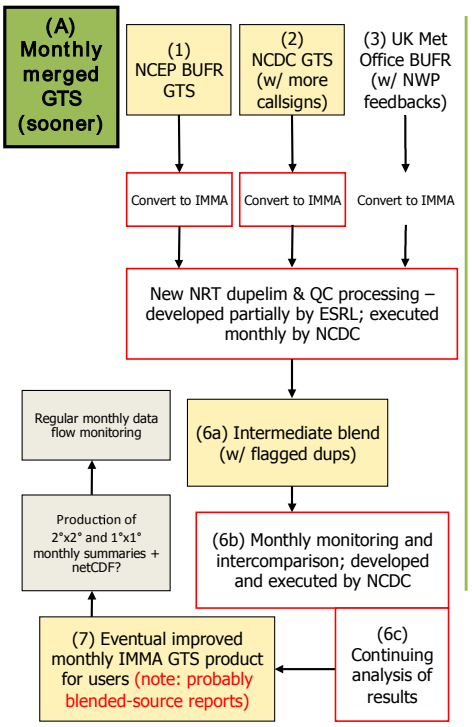
(16 April 2012 update note: titles following reconfigured, figures resized, etc. to resolve problems exporting to pdf)

(On this page) Previously under discussion among ICOADS project partners;

(Overleaf) Possible downsized NRT transition and improvement:



Possible Downsized ICOADS Modern Data Processing, Under Transition Plan



GTS data sources 1-3 are flowing regularly to NCDC/ NCAR:

- ~10% data receipt differences may exist between GTS sources (1) and (2) (ETMC-III tests revealed additional data differences, including w.r.t. to (3)).
- Translations of (1) and (2) need to be reconciled and re-validated.

Appendix E

Additional Non-Public Webpages Hosted at ESRL/PSD

(Note: available internally to PSD only, via the test web server)

(16 April 2012 update note: URLs are marked in grey because, as applicable, materials have already been migrated over to public locations)

Translation of the Deutscher Wetterdienst (DWD) archive data, and newly digitized data:
http://icoadstest.psd.esrl.noaa.gov/dwd.html
http://icoadstest.psd.esrl.noaa.gov/dwd2.html

Translation of the UK Marine Data Bank (MDB):
http://icoadstest.psd.esrl.noaa.gov/ukmdb.html

Contains codes used to generate "preliminary" IMMA and MSG (monthly summary statistics) from the NCEP BUFR data:
http://icoadstest.psd.esrl.noaa.gov/e-doc/other/transpec/ncep_bufr/

Appendix F

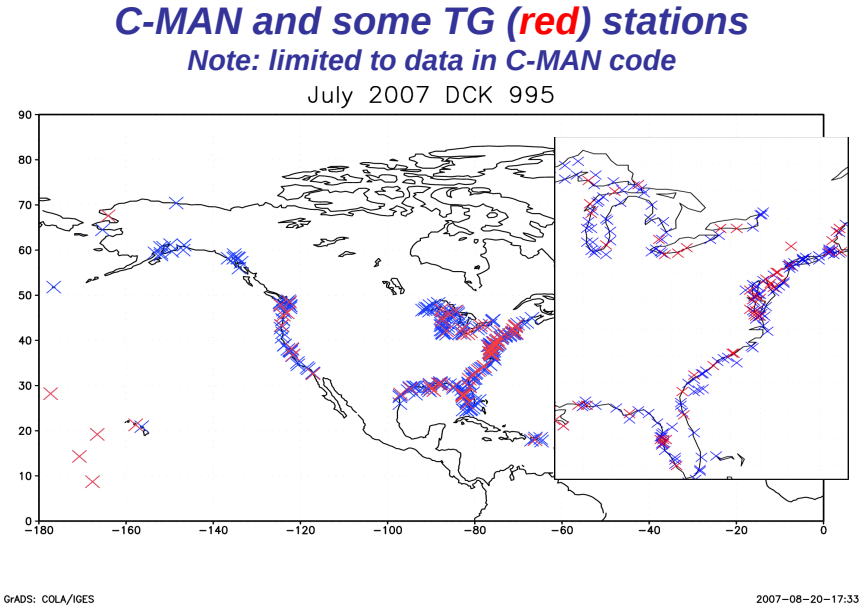
Tide Gauge (TG) Data Issues

a) Importance of TG data?

- SST, AT, winds, SLP, etc.
- >1/3 all GTS data gathered as part of "Marine"
- NCEP/NCDC each collect some unique stations

b) How about a shared US C-MAN/TG station list?

- Possibly already public from NCEP/NCO
  - But a data file might be better than Fortran data stmt.?
- Does NCDC have some additional stations?
  - NCDC C-MAN list only has LAT/LON to 0.1deg?
  - Systematic location differences exist with NCEP data
- Are differences between C-MAN/TG important?

c) Unlike other IDs circulated on the GTS (e.g., VOS and buoy) C-MAN and TG IDs not allocated/managed by JCOMM

- ✓ why not?