Data quality assurance and provenance tracking in ICOADS Release 3.0

Thomas Cram
Steven Worley, Zaihua Ji, and Doug Schuster
NCAR, Computational Information Systems Laboratory (CISL)

AGU Fall Meeting
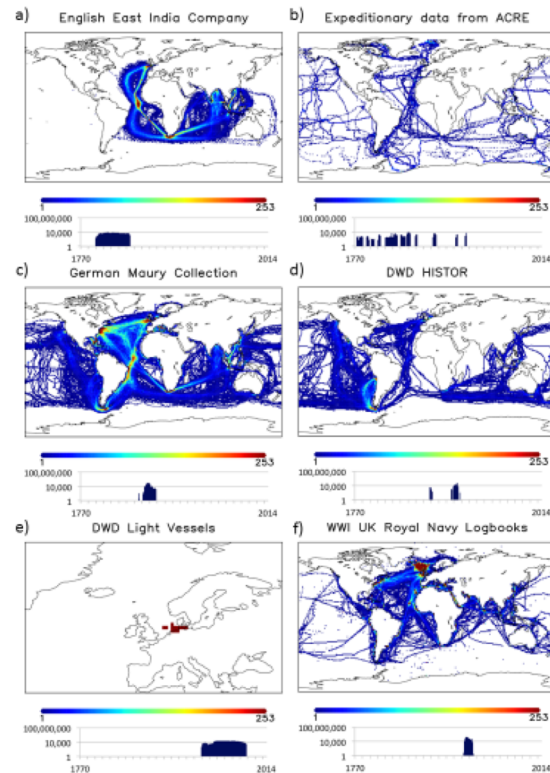IN33D-01
December 13, 2017

About the RDA

- NCAR's **R**esearch **D**ata **A**rchive
- Purpose
  - Support climate & weather research at NCAR and UCAR universities with reference datasets
- Collections
  - Ocean & atmospheric observations, analyses, reanalyses, operational NWP products
  - Established in 1960s
  - 700+ datasets, 8M files, 1.8 PB
  - 70+ datasets growing daily-monthly
- Free and open access
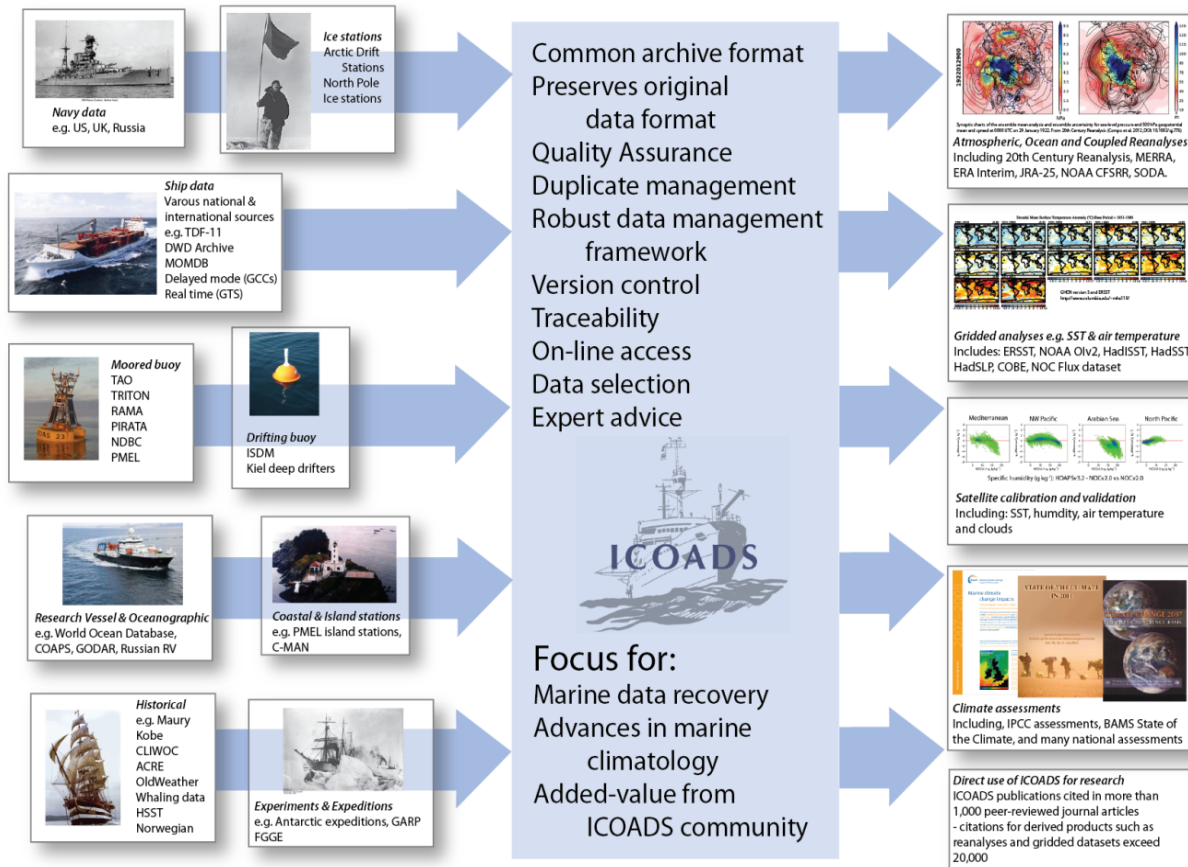- Worldwide usage
- Science educated staff

ICOADS overview

- **I**nternational **C**omprehensive **O**cean **A**tmosphere **D**ata **S**et

- World's most extensive collection of marine *in situ* data
- Foundational dataset for climate research
- **Release 1.0**
  - Published 1985
  - Temporal coverage 1854–1979
- **Release 3.0**
  - Published June 2017
  - Temporal coverage 1662–2014
  - NCEP/NCEI near real-time GTS extension (2015—ongoing)
- Access
  - NOAA: **icoads.noaa.gov**
  - NCAR: **rda.ucar.edu/datasets/ds548.0**

*Sample of recovered and digitized
data added to ICOADS R3.0*

Navy data
e.g. US, UK, Russia

Ice stations
Arctic Drift
Stations
North Pole
Ice stations

Ship data
Varous national &
international sources
e.g. TDF-11
DWD Archive
MOMDB
Delayed mode (GCCs)
Real time (GTS)

Moored buoy
TAO
TRITON
RAMA
PIRATA
NDBC
PMEL

Drifting buoy
ISDM
Kiel deep drifters

Research Vessel & Oceanographic
e.g. World Ocean Database,
COAPS, GODAR, Russian RV

Coastal & Island stations
e.g. PMEL island stations,
C-MAN

Historical
e.g. Maury
Kobe
CLIWOC
ACRE
OldWeather
Whaling data
HSST
Norwegian

Experiments & Expeditions
e.g. Antarctic expeditions, GARP
FGGE

Common archive format
Preserves original
data format
Quality Assurance
Duplicate management
Robust data management
framework
Version control
Traceability
On-line access
Data selection
Expert advice

ICOADS

Focus for:
Marine data recovery
Advances in marine
climatology
Added-value from
ICOADS community

Atmospheric, Ocean and Coupled Reanalyses
Including 20th Century Reanalysis, MERRA,
ERA Interim, JRA-25, NOAA CFSRR, SODA.

Gridded analyses e.g. SST & air temperature
Includes: ERSST, NOAA OIv2, HadISST, HadSST,
HadSLP, COBE, NOC Flux dataset

Satellite calibration and validation
Including: SST, humdity, air temperature
and clouds

Climate assessments
Including, IPCC assessments, BAMS State of
the Climate, and many national assessments

Direct use of ICOADS for research
ICOADS publications cited in more than
1,000 peer-reviewed journal articles
- citations for derived products such as
reanalyses and gridded datasets exceed
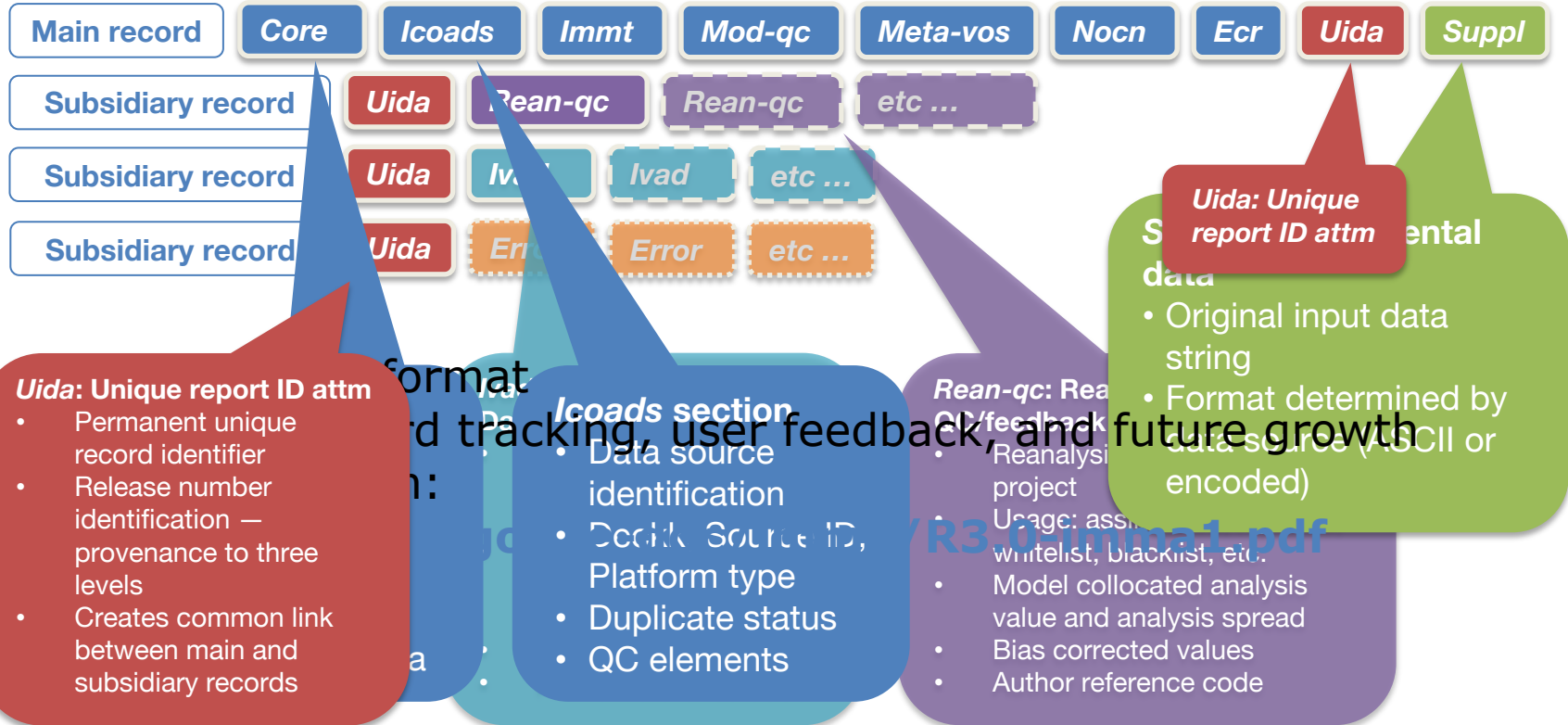20,000

New data sources in ICOADS R3.0
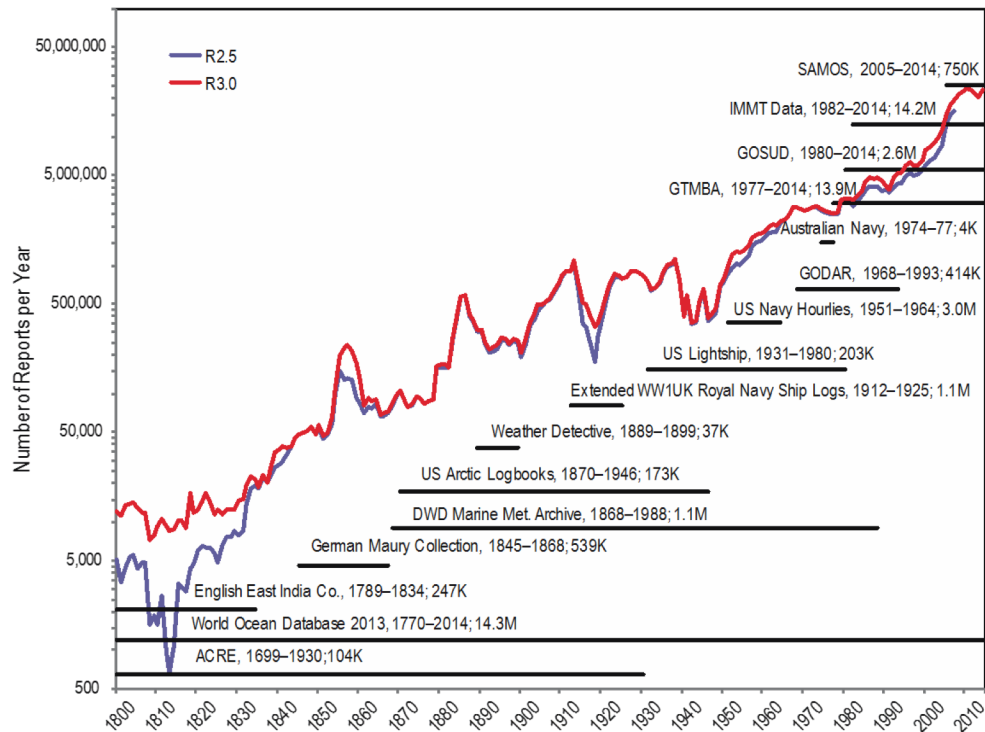
- **Red:** total number of reports in R3.0
- **Blue:** total number of reports in R2.5
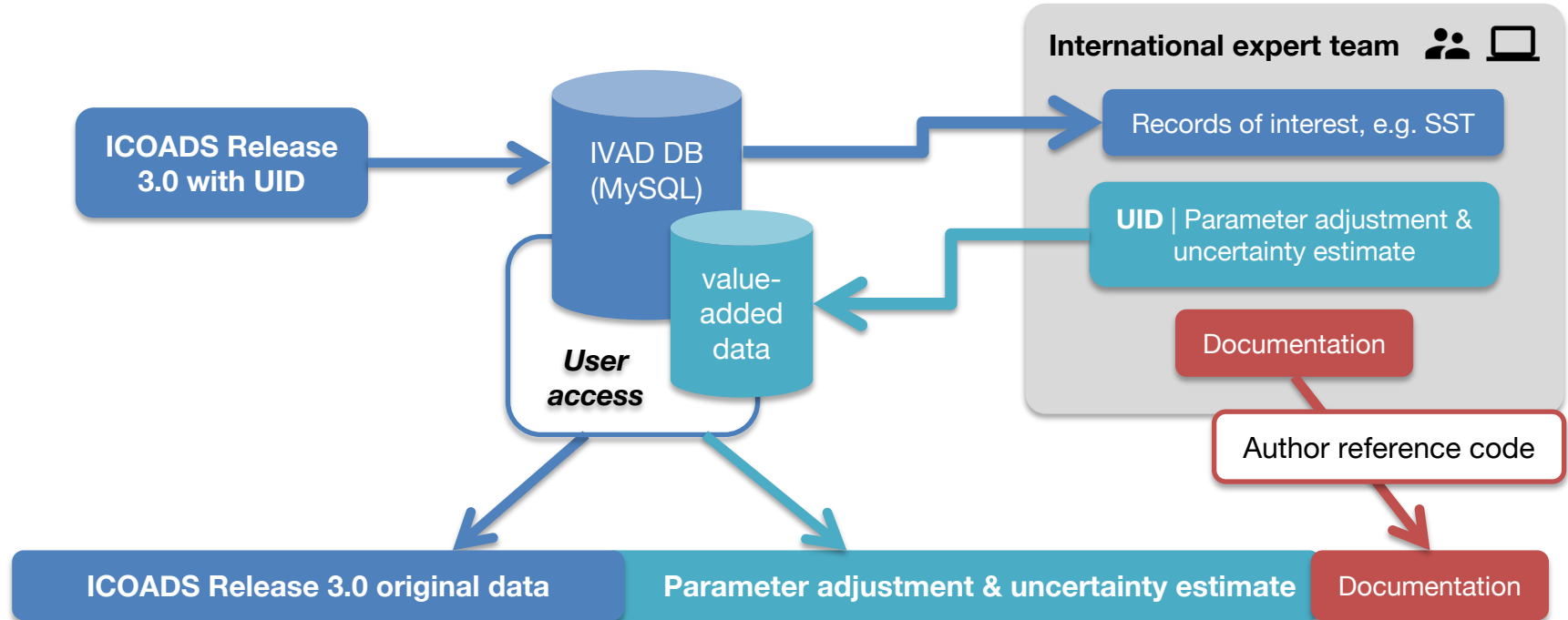- Horizontal black lines: primary data collections added to R3.0

455+ million records in total

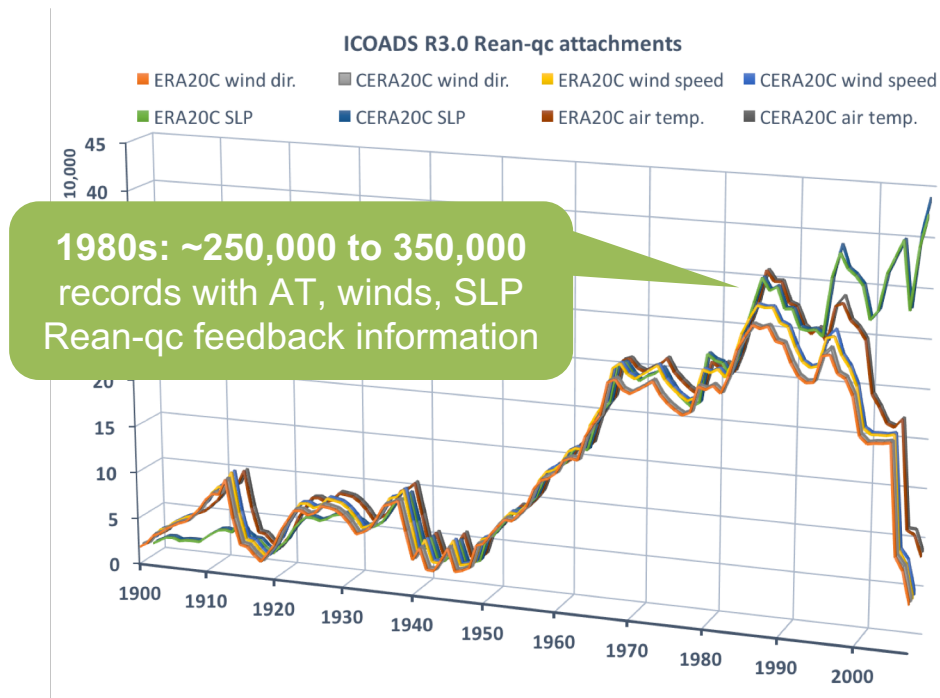# International Maritime Meteorological Archive (IMMA1) format

Main record	Core	Icoads	Immt	Mod-qc	Meta-vos	Nocn	Ecr	Uida	Suppl
Subsidiary record	Uida	Rean-qc	Rean-qc	etc …					
Subsidiary record	Uida	Ivad	Ivad	etc …					
Subsidiary record	Uida	Error	Error	etc …					

Uida: Unique report ID attm
- Permanent unique record identifier
- Release number identification — provenance to three levels
- Creates common link between main and subsidiary records

Icoads section
- Data source identification
- Deck, Source ID, Platform type
- Duplicate status
- QC elements

Rean-qc: Rean QC/feedback
- Reanalysis project
- Usage: assim, whitelist, blacklist, etc.
- Model collocated analysis value and analysis spread
- Bias corrected values
- Author reference code

Uida: Unique report ID attm

S... ...ental data
- Original input data string
- Format determined by ...data source (ASCII or encoded)

format
...rd tracking, user feedback, and future growth
...n:

...jo.../R3.0-imma1.pdf

ICOADS Value-added Database (IVAD)
Data flow, feedback management, user access

International expert team

ICOADS Release 3.0 with UID → IVAD DB (MySQL) → Records of interest, e.g. SST

UID | Parameter adjustment & uncertainty estimate

value-added data

Documentation

User access

Author reference code

ICOADS Release 3.0 original data | Parameter adjustment & uncertainty estimate | Documentation

Current *Ivad* & *Rean-qc* attachments

- *Ivad:* Value-added data
  - Berry et al. (2004) air temperature adjustment
  - Smith et al. (2015) Beaufort wind adjustment
- *Rean-qc*: Reanalysis/QC feedback
  - ECMWF ERA-20C: 20th century reanalysis
  - ECMWF CERA-20C: 20th century coupled reanalysis
- **Potential future contributions**
  - FSU Atlantic wind indices (S. Smith)
  - ECMWF ERA-5 feedback (?)

ICOADS R3.0 Rean-qc attachments

Legend:
- ERA20C wind dir.
- CERA20C wind dir.
- ERA20C wind speed
- CERA20C wind speed
- ERA20C SLP
- CERA20C SLP
- ERA20C air temp.
- CERA20C air temp.

1980s: ~250,000 to 350,000 records with AT, winds, SLP Rean-qc feedback information

Data quality assurance

Pre-defined trimming levels (based on 3 historical periods)

- — Standard
    - • 3.5σ threshold
    - • Ship obs. only
- — Enhanced
    - • 4.5σ threshold
    - • Ships + buoys + others
- — Custom
    - • User-defined filters

○ Standard Filtering
● Enhanced Filtering
○ Make your own Filtering Selection

day/night options	☑ day & night obs.	☐ night obs. only	☐ day obs. only	
platform type options	☐ ships obs. only	☑ ships + buoys + others		
source exclusion flags	☑ used	☐ ignored		
NCDC QC flags	☑ used	☐ ignored		
outlier trimming level	☐ 2.8 sigma	☐ 3.5 sigma	☑ 4.5 sigma	☐ untrimmed
if trimming flag=11	☐ reject obs.	☑ use obs.		☐ no effect

Trimming selection interface at rda.ucar.edu/datasets/ds548.0

Summary

- IMMA1 format is designed to support:
  - Record tracking (data provenance)
    - Unique record identification (UID)
  - User feedback
    - Value added data (IVAD)
    - Reanalysis feedback
  - Future growth
    - Encourages further data development & community engagement
    - Enriches the data beyond the scope of primary marine observations
- Completing the data life cycle

Contact us

🌐 *rda.ucar.edu*

✉️ *rdahelp@ucar.edu*

🐦 **@NCAR_RDA**

f **@NCAR.RDA**

🅱️ *ncarrda.blogspot.com*

Research Data Archive

Contributors: E. Freeman (NOAA/NCEI/ERT, Inc.)
A. Anderson (DWD)
W. Angel (NCEI)
P. Brohan (UK Met Office)
L. Gates (DWD)
E.C. Kent (NOC)
C. Liu (NOAA/NCEI/ERT, Inc.)
S. Smith (FSU)
H. Zhang (NOAA/NCEI)