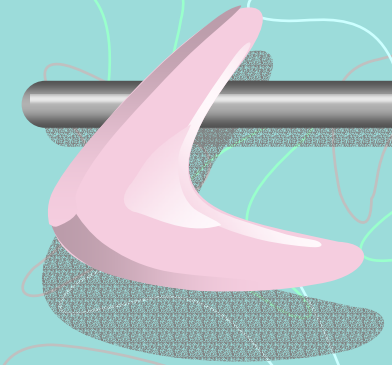
TOWARDS A NEW KNOWLEDGE OF GLOBAL CLIMATE CHANGES: METEOROLOGICAL DATA ARCHIVING AND PROCESSING ASPECTS



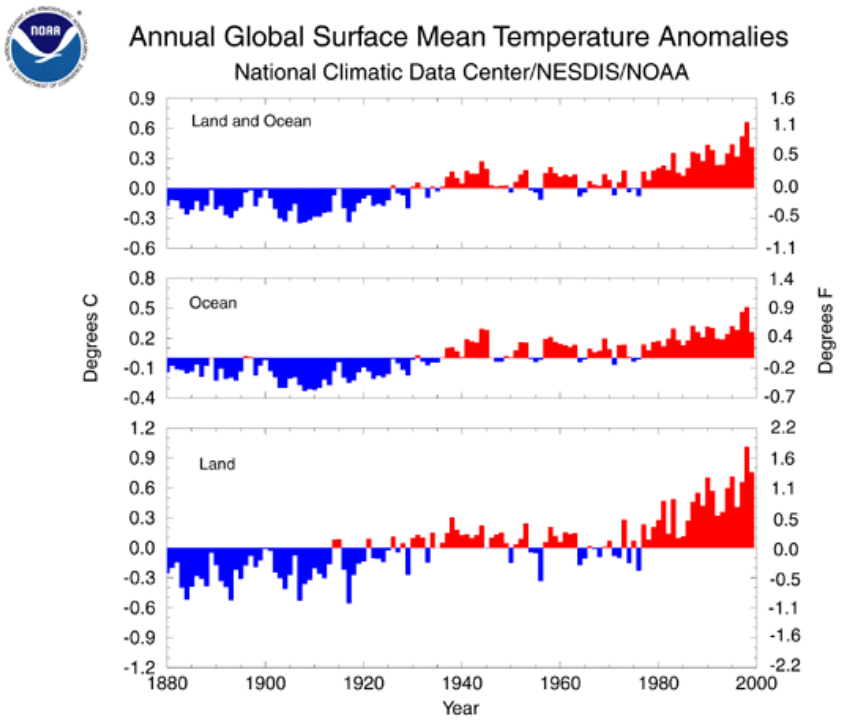
Dr. Alexander Sterin,

RIHMI-WDC, Russia

Russian Research Institute for Hydrometeorological Information – World Data Center (RIHMI-WDC)

- RIHMI-WDC is located in Obninsk, 100 km from Moscow
- RIHMI-WDC hosts 4 World Data Centers: Meteorology, Oceanography, Rockets & Satellites, Earth Rotation
- RIHMI-WDC has one of the most wide collections of global meteorological, aerological, hydrological, oceanographical data in the world
- The total volume of the digitized data in RIHMI-WDC is about 2 TBytes
- The total volume of hardcopy undigitized data at RIHMI-WDC is questionable

IPCC 2000 Third Assessment Report declared that global warming occurs, and over the 20th Century, the increase of global average surface T has been 0.6+/-0.2 Deg. C



- To obtain the estimates of such climatic trends in the global temperature, one needs to process the total of about 200 GBytes of primary observational data obtained for land surface, sea surface, atmosphere

Moving towards new knowledge of climate variations, we need to resolve the following problems:

- *Environmental observational Data collection and transfer (should not be mentioned here)*
- *Data archiving and preservation*
- *Data quality assurance and control*
- *Methodology of Data processing*
- *Analysis of Data, getting and presenting new Information on Climate Changes*

Data archiving and preservation

- The equal absolutely identical copies of data archives must be located on diverse media in diverse places (even in the diverse buildings)
- Periodical transfer to new media is required
- Old media and old media drives must not be lost
- Currently RIHMI-WDC collection is transferred from obsolete 9-track tape media to new tape library media (SuperDLT tapes)
- The total amount of 9-track tapes is about 60,000
- Such transfer is not a simple one-moment action

Data archiving and preservation: the transfer to new media

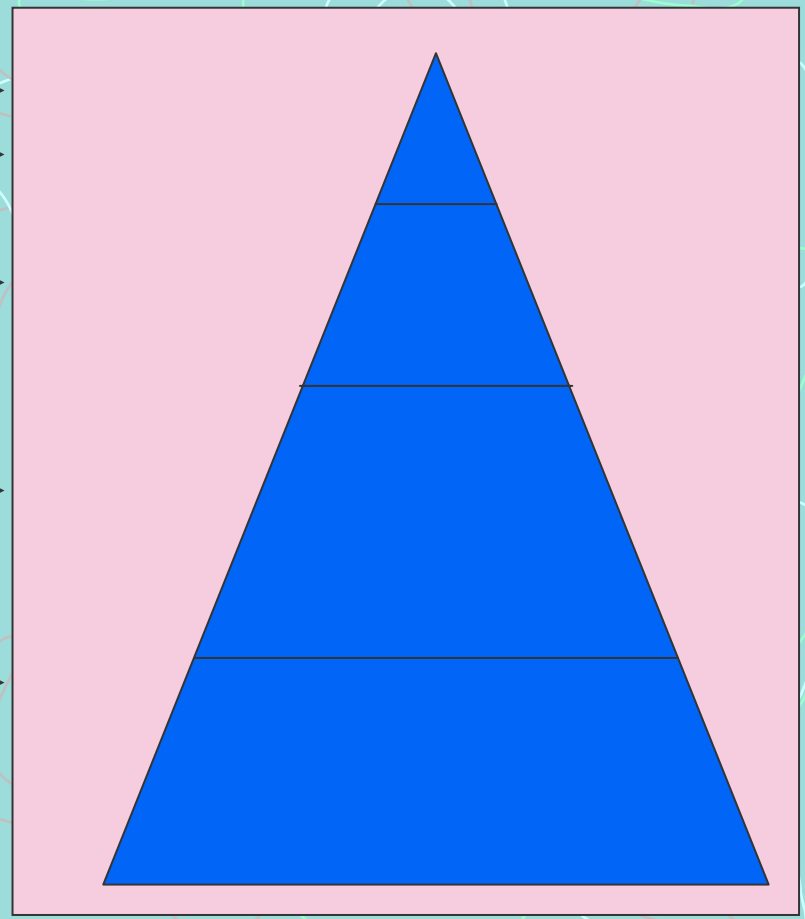
- New solutions –technical, scientific, technological, financial, organizational – are needed
- Special solutions on how to re-arrange Data portions are needed
- Catalogue of Data must be created in parallel with the copying process
- Metadata files must be created
- Data conditioning (integrity check) process must go in parallel
- All conditioning (integrity check) process must be put in protocol. Protocol will be preserved
- A step back (or few steps back) must be possible at any point of the process

Data Quality Control (QC): the basics

- *For QC of meteorological Data, complex solutions that are based on physical, meteorological, technological, linguistic considerations are needed*
- *QC procedures must not change any values of Data – only QC flags must be attributed to values, so that step back should be possible at any point*
- *QC algorithms are not magic – for meteorological data, they often decline from QC solutions and flagging (time series with gaps, sparse network of stations, etc.)*

Methodology of Data Processing: The Pyramid of Climate Products

- Global climate generalizations (separate figures)
- Highly integrated climate data derivatives (miserable volume)
- Climate data derivatives based on inputs of next lower level derivatives (small volumes)
- Climate data derivatives based on inputs of observational Data (modest volumes)
- Next level: observational Data after QC (huge volume)
- Lowest level of climate Data products: raw observational Data (huge Volume)



Methodology of Data Processing: samples of multi-level structure of climate Data products

- *CARDS (Comprehensive Aerological Reference Data Set) – a joint project of NCDC/NOAA, USA, and RIHMI-WDC, Russia*
- *MONADS –CARDS Derivative – MONthly Aerological Data Set – Monthly statistics for climate research, calculated from CARDS observational data*
- *Further aerological climate products are intended to be calculated from MONADS, not from CARDS observational Data*

Methodology of Data Processing: samples of multi-level structure of climate Data products

- *COADS – Comprehensive Ocean – Atmosphere Data Set*
- *COADS produced a lot of climate derivatives based on observational data*
- *REANALYSIS Projects: the main goal is to produce meteorological fields for further climate studies and modelling*
- *REANALYSIS Project of NCAR/NCEP (USA) produced a lot of derivatives. Multi-level structure of these derivatives is strongly evident*

Methodology of Data Processing: cycles of Data processing

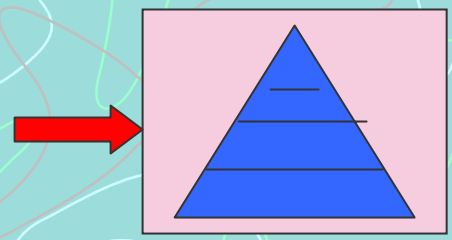
- Based on previous experience and on better understanding of climate processes, several cycles are needed for data archiving and processing within large climate-related data projects
- Technical progress requires to repeat cycles of climate data processing & archiving
- All stages of archiving must be available: you need to reutilize data collections, to update them and to do the recalculations of climate products
- Normally, these cycles must be repeated once or twice per decade

Analysis of Data

- Lot of analysis methods must be applied to data for climate studies (*elementary statistics, data mining, regression, multidimensional analysis, cluster analysis, time series analysis, spectral and wavelet analysis, etc*.)
- Appropriate presentation instruments (*graphs, maps, reports, etc*) must be applied
- Spatial and temporal inhomogeneities are typical for environmental data
- Special analyses enable to detect and to adjust these inhomegeneities
- Are adjusted Data better than unadjusted Data?

Analysis of Data: which software instrumentation should be used?

- *The Software must:*
- enable access to archived data in their native formats
- support operations on all levels of the pyramid of climate products
- Have powerful analytical capacity
- Offer numeric graphical capabilities, including mapping
- Special original software ?
- SAS ® ?



Some concluding remarks

- Getting new knowledge on climate changes needs processing of huge volumes of observational Data from various platforms worldwide
- These Data are located in various places, on various media and in various formats – you need to be ready to overcome this diversity!
- These Data, both raw and QC'd, need to be archived, and transfer to new archival media must go on, as soon as new technical and technological solutions appear

Some concluding remarks

- The multi-level structure of climate Data products enables to avoid repeated processing of huge volumes of observational Data in most situations, but not always
- The multi-cycle processing of observational Data for climate studies is natural in global climate projects
- To improve our knowledge of climate changes, you need to do numerous inter-comparisons of Data from various sources
- However, our knowledge of climate changes remains incomplete and uncertain!

Thank you!