

## Data Quality Control and Archiving System for AWCI Data Center

Eiji Ikoma\*, Hiroko Kinutani\*,  
Masaki Yasukawa, Toshihiro Nemoto\*\*  
and  
Masaru Kitsuregawa\*\*, Toshio Koike\*\*,\*\*\*

The University of Tokyo

\*Earth Observation Data Integration and Fusion Research Initiative, UT

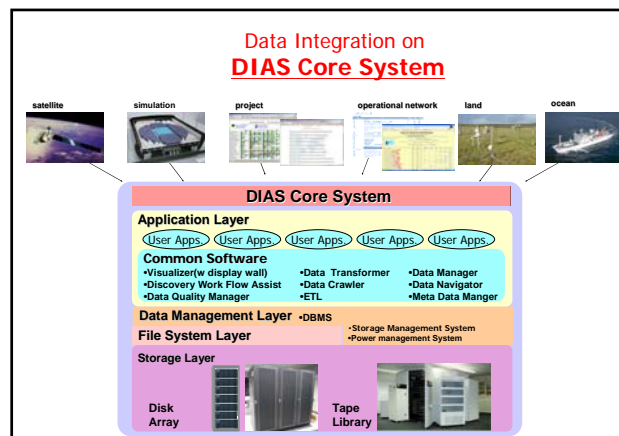
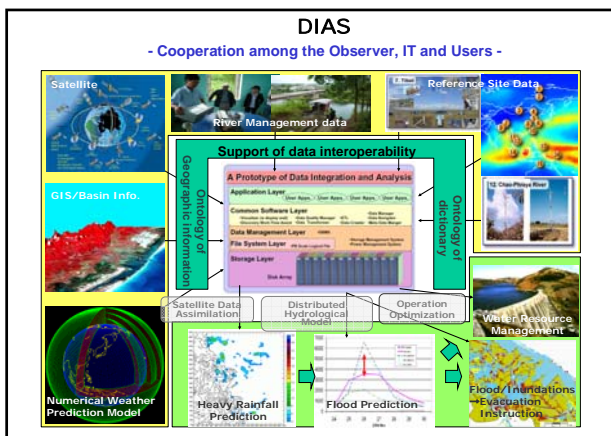
\*\*Institute of Industrial Science, UT

\*\*\*Department of Civil Engineering, UT

## Outline

1. Outline of Data Upload, Quality Control, and Meta-Data Registration System on *DIAS*.
2. Introduction of **Data Upload System**
3. Introduction of **Data Quality Control System**

→ cont. Dr. Kinutani's presentation



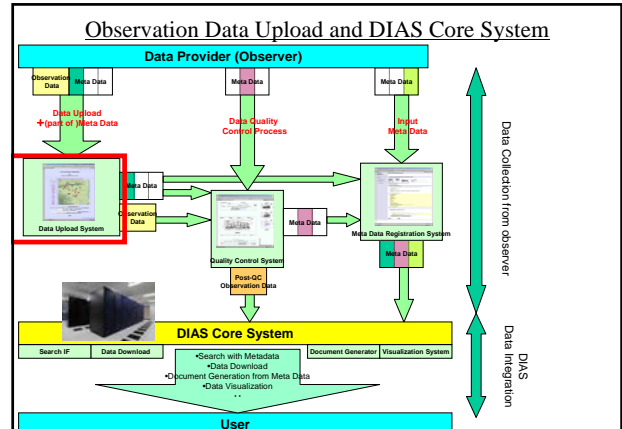
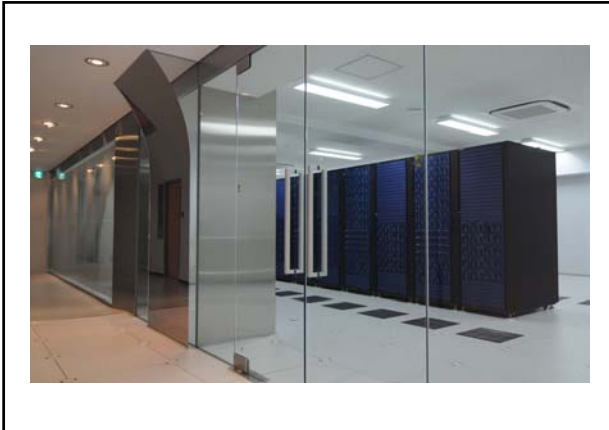
## Storage System Structure

### DIAS Storage System

- Connect to Dual 8-way Itanium2 Server
- 5-subsystem(Storage)
- Read 2.4GB/s/subsystem, total 12GB/s
- ≒ 1PB capacity HDD
- Advanced power management
  - Automatically turn off the power of idle part
  - Automatically turn on the power when accessed

### New Installation around Petabyte-scale Storage at Institute of Industrial Science, U-Tokyo





## Observation Data Upload System

Eiji Ikoma

Katsunori Tamagawa, Hiroko Kinutani,  
Tetsu Ohta, Toshio Koike, Masaru Kitsuregawa

## On-line Data Upload

- Observers have sent their own data to data administrators with the means of e-mail or mail before.
- However, by these methods, there were lots of problem, like the point that the file format and meta information are not unified, the point which requires much time and effort to send the data for observers, and also processing take lots of time, etc..
- So, we have started to develop on-line data upload system for AWCI data, which is in cooperation with Data Quality Control System, Meta-Data Registration System, and Data Archiving System.

## Data Upload System

- Observers can upload observation data and input some Metadata on Web Interface consisted of 4 steps.
- On each step, observers need to input some information about the data.
- Easy Operation and Quick Response.
- This system has some function which **reduce** the complicatedness of upload process

## Login Page



- Username and Password are required.
- Each observation site manager has its own (unique) username and password.

### STEP1



- Observation Point(Map/List)
- Time Period
- Data Interval
- Timezone
- Description (optional)
- Num. of observed elements

### STEP2



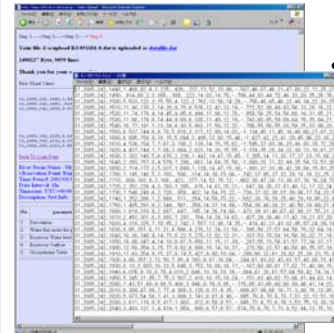
- Observation Data
    - Choose from pulldown menu
  - Sensor height
  - Orientation(op.)
  - Unit
  - Missing value
  - Description(op.)
1. Copy from No.1 to all
  2. Unit = Input Automatically when you choose observation data
  3. Copy from former inputted data
  4. Modify the num of observation data
  5. Upload from prepared csv file

### STEP3



- Upload observation Data(File).
- Confirmation of metadata inputted at STEP1,2.

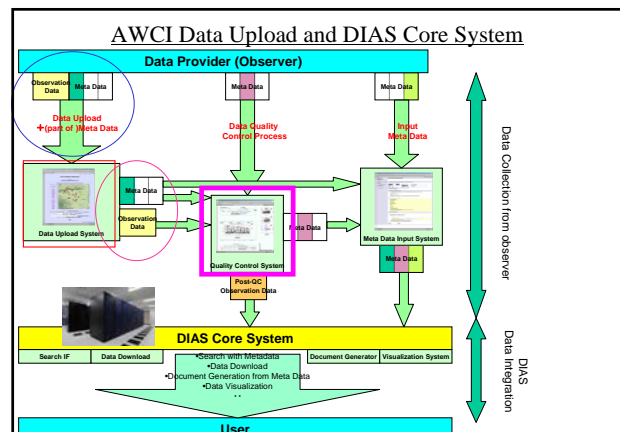
### STEP 4



- Confirmation of
  - local path of uploaded file
  - contents of the file (first/last 3lines and all lines when you require)
  - All metadata inputted at STEP1,2,3

### After STEP 4

- Our system send the confirmation message to observer by e-mail.
- Inputted metadata are stored in our Upload system --- Observer can use at next time.
- Observation data is loaded to Quality Control System



## Data Quality Control(QC) System

**Eiji Ikoma, Katsunori Tamagawa,**  
Tetsu Ohta, Kenji Taniguchi,  
Toshio Koike, Masaru Kitsuregawa

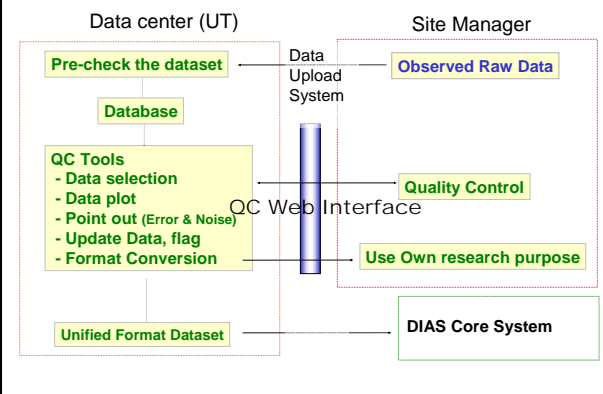
## Our QC System

- First version of our QC System was developed for CEOP Data in 2004.
- Ver.1(2004-2005) for → Ver.2(2005-2006) → Ver.3(2007-)
- 13site(Ver.1)→ 25site(Ver.2)→Ver.3
- We are running QC-V3 system for AWCI Observation Data.

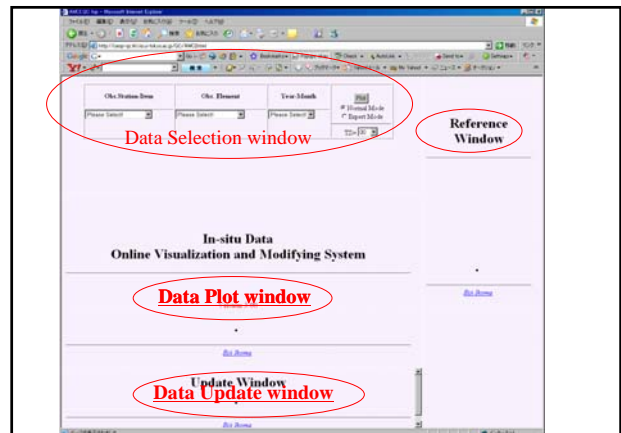
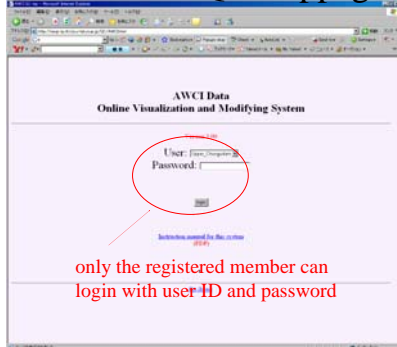
## Features of our QC system

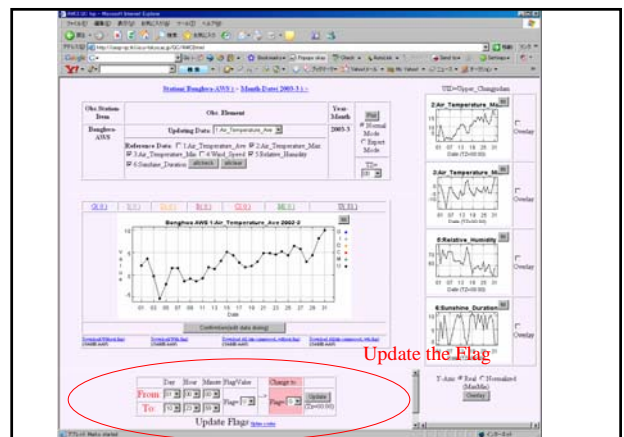
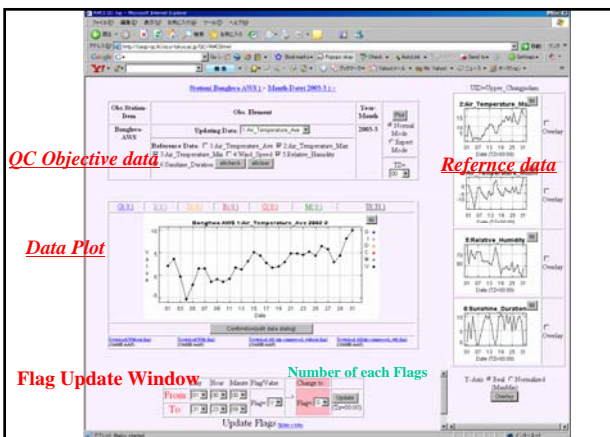
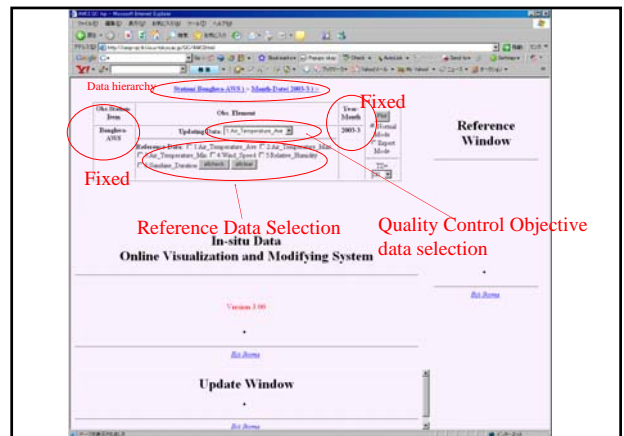
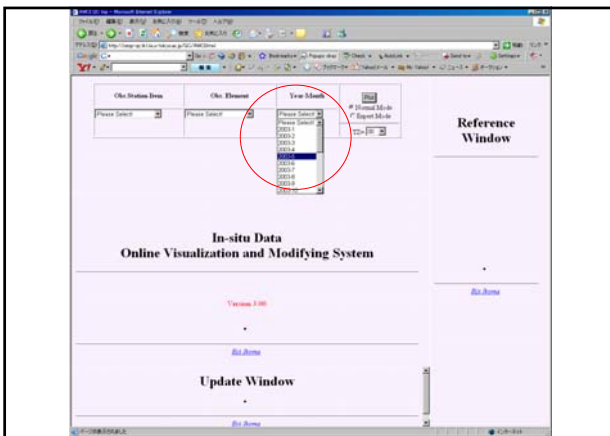
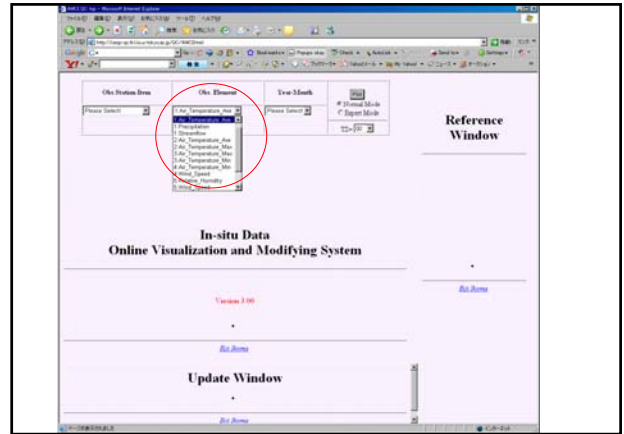
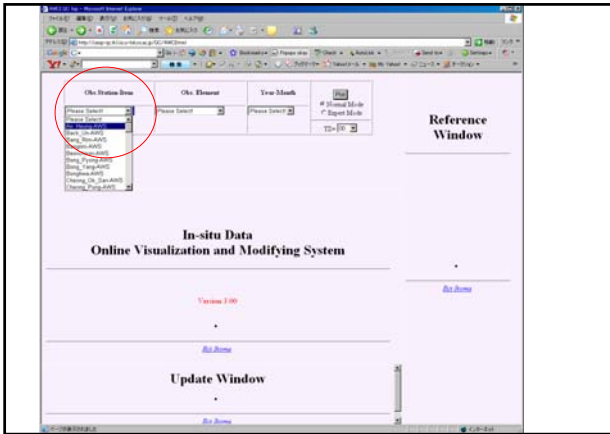
- Web-based UI ( required only Web browser)
- Easy-to-use and light operation
- Data management mechanism for each user authority
- Post-QC Data download support system
- Progress Management system for Data Manager

## Outline of QC Process



## AWCI Data-QC Toppage





**Flag Definitions**

- G**: Good
- I**: Interpolated
- D**: Dubious/Questionable
- B**: Bad
- A**: Abnormal value
- M**: Missing
- U**: Unchecked

**Number of each Flags**

**Flag Updated data**

**Click on the point**

**Updating Data**

**Updating Data**

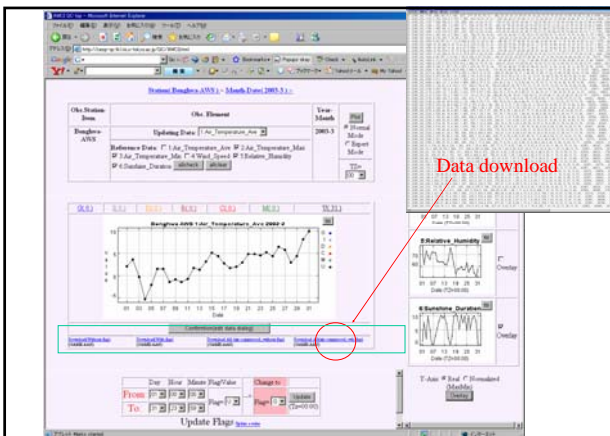
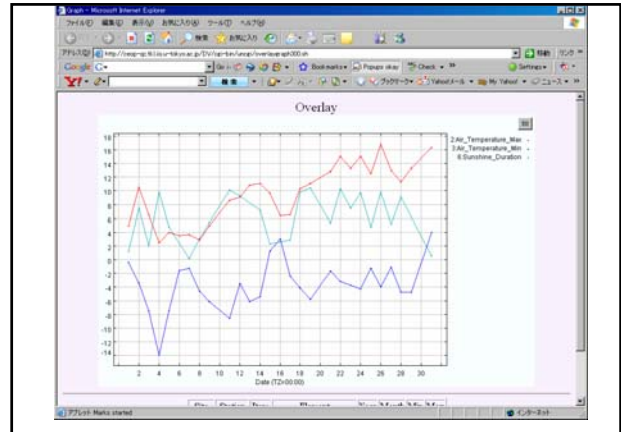
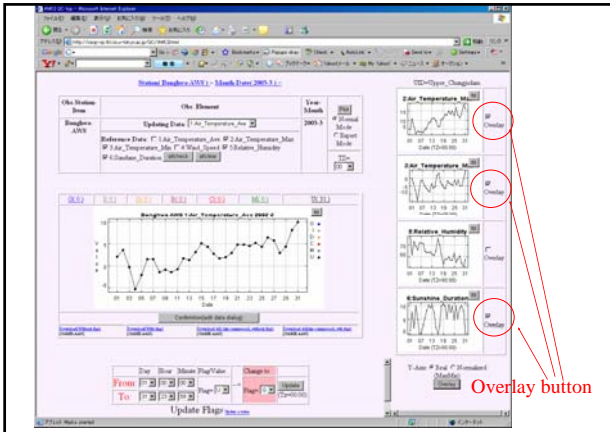
### Confirmation for update

**Update List(confirm)**

Update DATE-TIME	Original	Modified
2003-3-17 00:00:00	value=0.7 flag=U	value=0.7 flag=C

**Update**

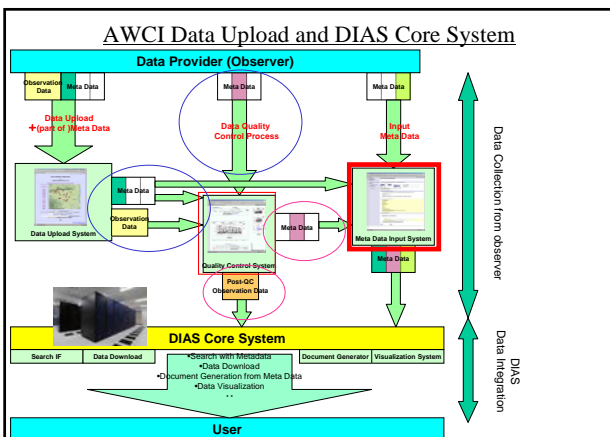
**Updated Data**



### GEOSS/AWCI Data Status

Country	Basin Name	Basin Info.	Raw Data Unload	Complete DB	Quality Control	Meta Data Unload	Remarks
1	Bangladesh	Meghna		△			
2	Bhutan	Punatsangchhu					
3	Cambodia	Sangker	10/30	△			
4	India	Seonath	7/22	○			
5	Indonesia	Mamberamo					
6	Japan	Tome	10/30	○			
7	Korea	Upper Chungju-dam	8/05	10/02	11/02		
8	Lao PDR	Sebangfai					
9	Malaysia	Langat					
10	Mongolia	Selbe	7/22				
11	Myanmar	Shwepyin					
12	Nepal	Bagmati					
13	Pakistan	Swat	7/22				
14	Philippines	Pampanga	8/05	○			
15	Sri Lanka	Kalu Ganga	8/05				
16	Thailand	Mae Wang	8/05				
17	Uzbekistan	Chirchik-Okhangaran	8/05				
18	Vietnam	Huong	7/22	○			

MMIDD: Provided to AWCI Data Center  
○: Full Data Provided by Offline  
△: Partial Data Provided by Offline



- アップロード+QC+メタデータ登録
- 如何にデータをアップさせるか、如何にメタデータを入れてもらうか。
- 入れる気にさせるポイントを散りばめて、なんかキラキラしたものを。
- 進捗を見せてプレッシャーを。

## 話の流れ

- 最初に全体構成(いこま)
- 各サイトの進捗(いこま)
- アップロード(いこま)
- QC(いこま)
- メタ入力(絹谷)、デモ
- 出力結果、おいしいポイント(絹谷)

## おいしいポイント？

- PDFドキュメント(=提出が楽?)
- 入力例から予想される使い方スライド(玉川?)  
←ある地点の周辺のデータが検索できる!など..
- QCにも役に立つ!参照がしやすい。

## お願い

- 良い入力例、その結果の出力、それで実現できる(であろう)利用例を考えて下さい