

# Practices of DAR metadata at GISC/DCPC Tokyo

TOYODA Eizi

toyoda@npd.kishou.go.jp

2012-04-16 TT-ApMD-1

# Timeline

Jun 2011

Cg-XVI designated GISC/DCPC Tokyo

Aug 2011

JMA declared GISC/DCPC operational

Feb 2012

WMO declared GISC/DCPC operational

# Three Types of Data

	From NC	From DCPC
On Volume C1	SYNOP, TEMP	NWP GRIBs
Not on Volume C1	?	GAW WDCGG

# Before coming to specifics ....

- Sometimes people say:  
“the more information, the better metadata”
- Is that true?

# Trade-off

- Too few information
  - Maybe problem in some use
- Too much information
  - Hard to understand
  - Less likely for people to join & to keep maintenance

# We must seek balance

- Mandatory elements limited to info with important use (application)
- Keep others optional

# What's application in discovery MD?

- Search
- Others
  - For GTS bulletins:  
Vol C1 used in monitoring, data management, ...
  - For other data, sometimes there are existing catalogues of data

# My conservative approach

- First guess:  
elements of traditional catalogue
- First correction to add:  
search keys necessary to identify dataset

# Three Types of Data (again)

	From NC	From DCPC
On Volume C1	SYNOP, TEMP	NWP GRIBs
Not on Volume C1	?	<b>GAW</b> <b>WDCGG</b>

Volume C1 is for later session ...

# World Data Centre of Greenhouse Gases

- Under Global Atmospheric Watch Programme
  - Collecting data worldwide
  - Servicing on the web
- Traditional catalogue exists
  - Latest version is XML
- What I did
  - An XSLT to convert it into ISO 19115/139
  - Links to data added

# WDCGG metadata structure

- <wdcgg>
  - <station>
    - <name>, <id>, <lat>, <lon> ...
    - <parameter>
      - <name>, <start\_date>, <end\_date>, ...
      - <contributor>
        - <organization>, <country>, ...
      - <contact\_point>
        - <name>, <organization>, ...

# Three Types of Data (again)

	From NC	From DCPC
On Volume C1	SYNOP, TEMP	<b>NWP GRIBs</b>
Not on Volume C1	?	GAW WDCGG

# Grid data from Numerical Weather Prediction

- Single model run → 20,000 messages of GRIB
  - Splitted by:  
*resolution, bounding box, vertical level, forecast time, physical parameter, ....*
- Volume C1
  - Painful to maintain
  - Not so great for users
  - Because *keys for identification* are packed in a free-form text field “Contents” or missing

# What I did for NWP GRIBs

- Per-bulletin MD records (22000)
  - Identification keys added to VolC1 columns
  - resolution, bounding box, vertical level, forecast time, physical parameter, ....
- Aggregated MD records (15)
  - Common: model, resolution, & bounding box
  - Children: hundreds of per-bulletin MD's
    - Link by “parentIdentifier” (URN fileIdentifier)
    - URL to parnt metadata added

# Summary

- Respect traditional catalogus
  - Never re-invent wheels!
- Search key is important
- Group of too many records should have “parent” (aggregate) metadata record