Metadata technology in WMO Information System (WIS)

TOYODA Eizi
toyoda.eizi@gmail.com

2011-05-26
Japan Geoscience Union Meeting

Topics of the day

- Metadata format
- User interface
- Inter-organisational metadata exchange
- Next step

What's WIS?

- WIS: WMO Information System
  - WMO: World Meteorological Organisation
- Continues & enhances GTS
  - Realtime network for operational meteorology
  - Legacy long before Internet
- Infrastructure for all WMO programs
  - Now efforts concentrates on data catalogue

Organisational structure

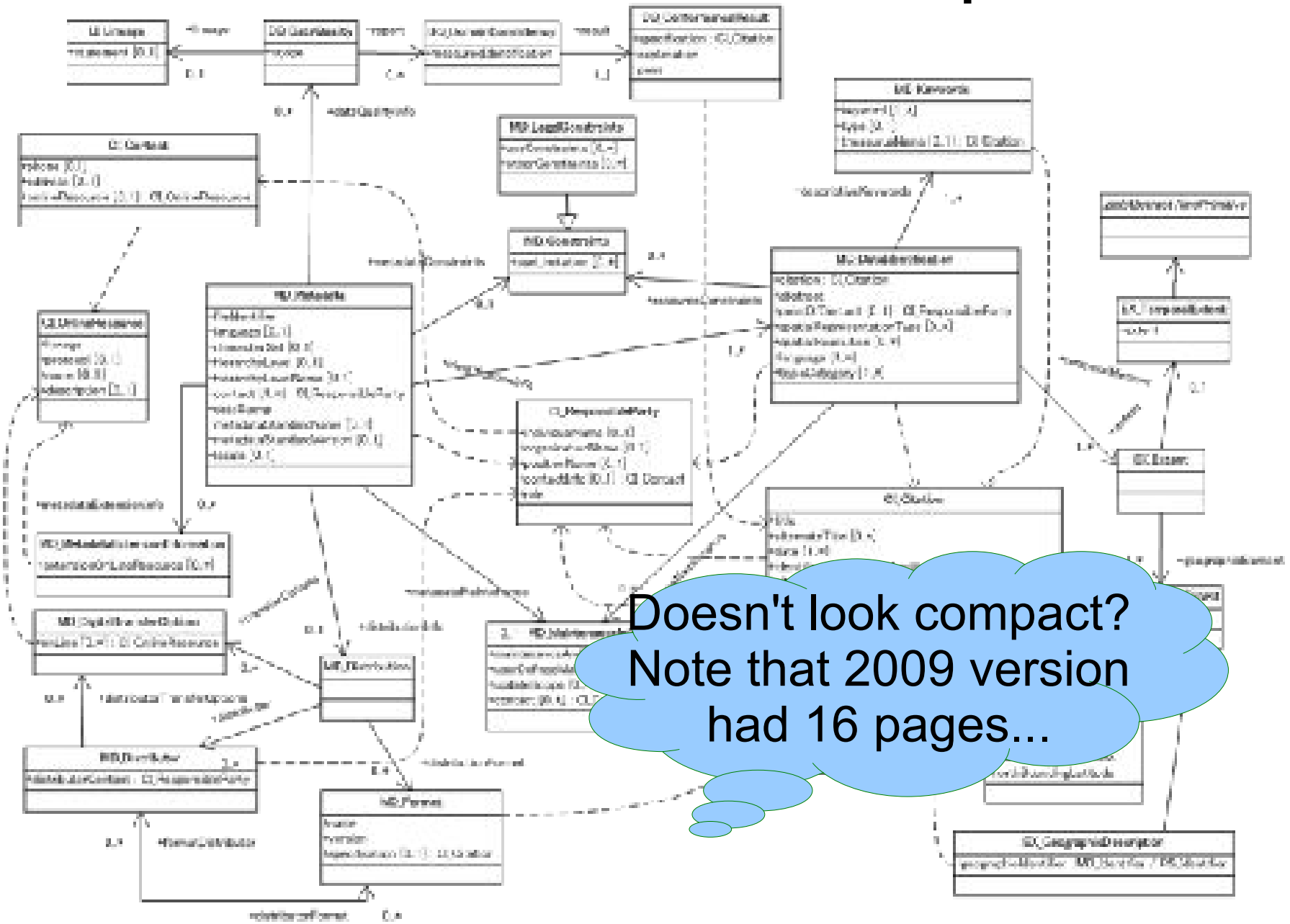
- GISC: global information system centre
  - information catalogue of entire WIS
  - global distribution information on web
  - network management
- DCPC: data collection & product centre
  - regional or programme-wide activities
- NC: national centre
  - national activities

Catalgue is
key feature of
central WIS centres

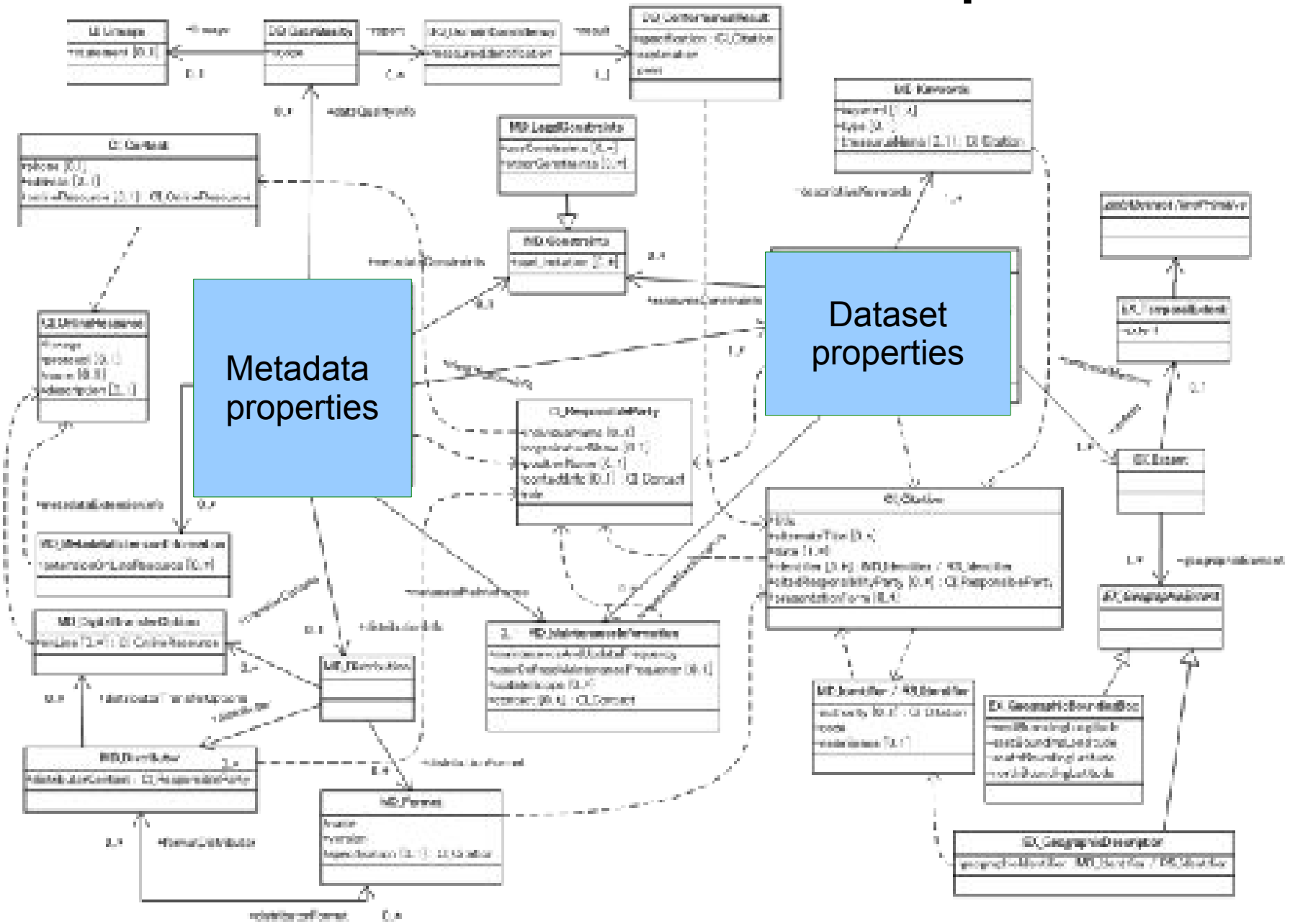
Finally metadata structure discussed

- WMO adopts ISO 19115 (2003)
  - Firstly extensions to ISO was discussed
  - Goal is complete description of data ...
- WMO Core Metadata Profile v1.1 (2009)
  - All extensions thrown away
    - Cannot validate against XML Schema
    - Cannot understood by people out of community
- Current efforts on narrowing-down
  - What is "core elements" which everybody must fill?

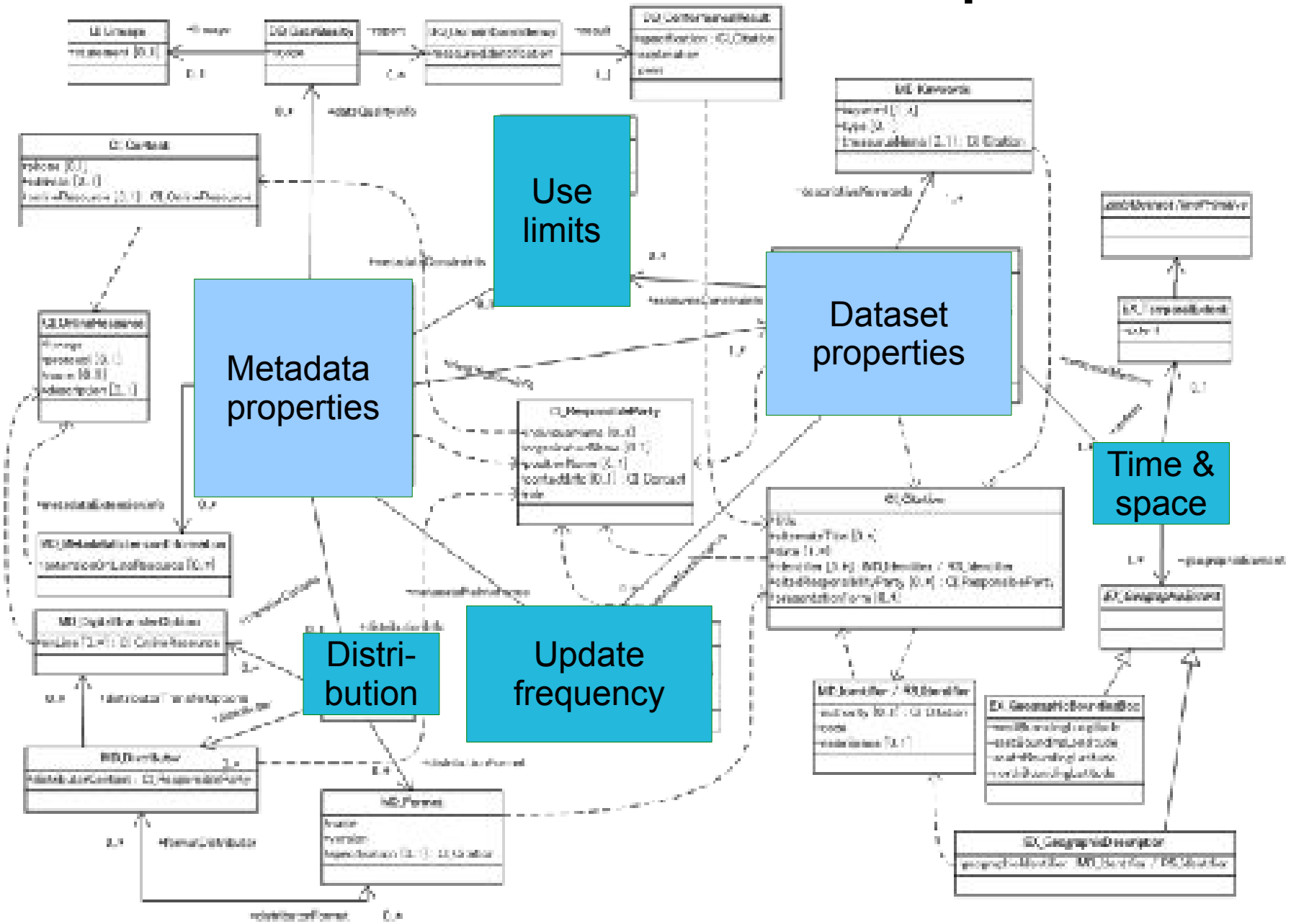
Look! Now it's so compact!



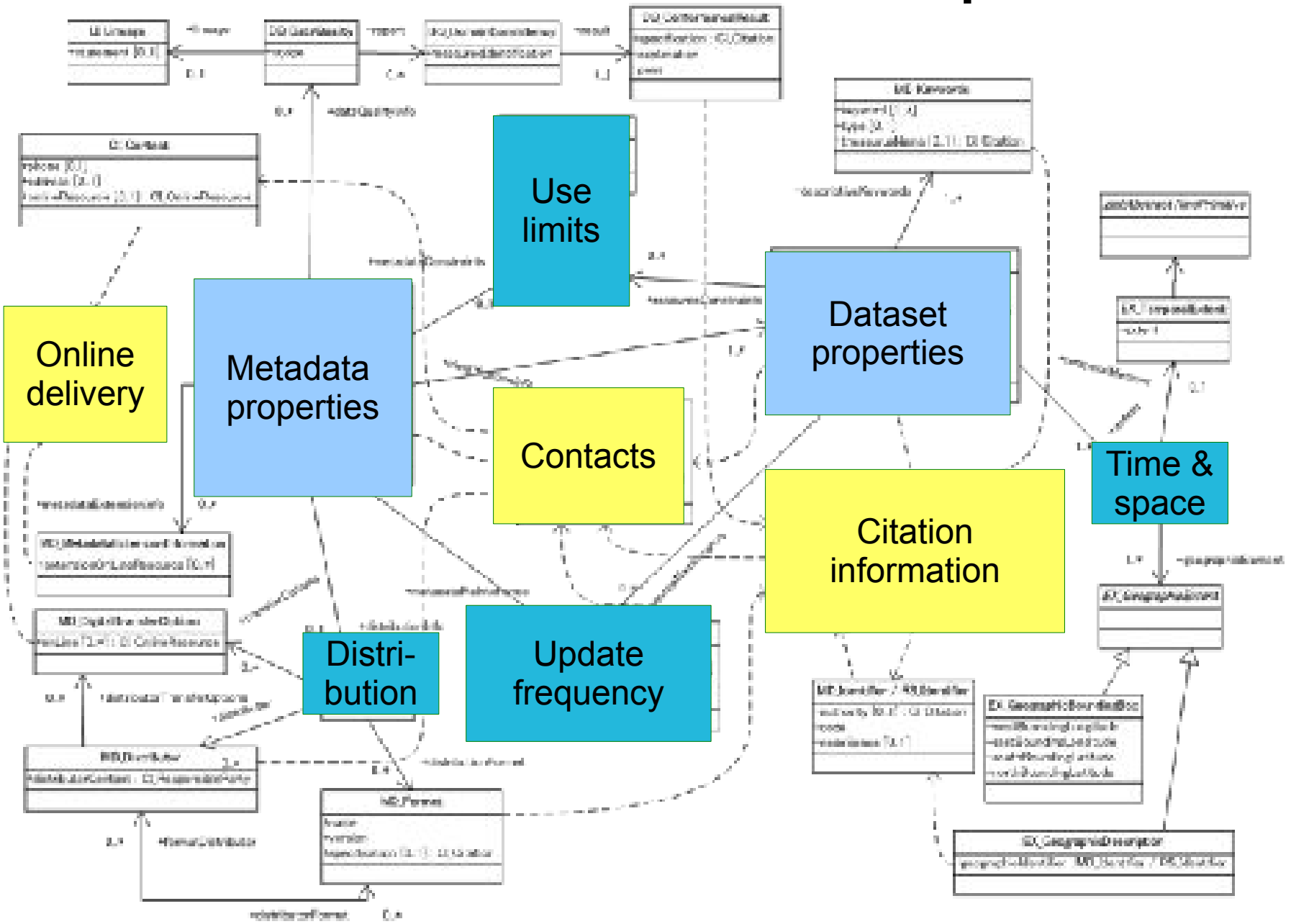
Look! Now it's so compact!



Look! Now it's so compact!



Look! Now it's so compact!



Size of record schema

- WMO Profile gets compact, but still has 280 simple-typed elements
  - Who can remember all?
  - But only 14 are mandatory
- Practical record schema cannot exceed 30 elements
  - Typical legacy catalogues: 10 ~ 30 example

Efficient input toolig

- Online editor on web
- MS Excel
  - We did validation by VBA macros before
  - But stopped
    - Computer-friendly format will help automated submission

User interface

SRU protocol

- Developed in librarian community
  - Aim: to replace ISO 23950 = ANSI Z39.50
- Technical feature similar to OGC WMS
  - Request: CGI-like query (or SOAP)
  - Result: XML including metadata records

SRU good for mashup

- Mashup: third-party use of web service, to create added-value site
  - E.g. twitter API, Google search box
- SRU result is XML: easy to use
  - Caveat: JavaScript XSS issue

SRU Index

- Like "title" in <<title = "temperature">>
- Parallel search requires standardised index
- WMO trying to start mapping between "19115 to SRU index"

SRU indices in (draft) WMO manual

STRING

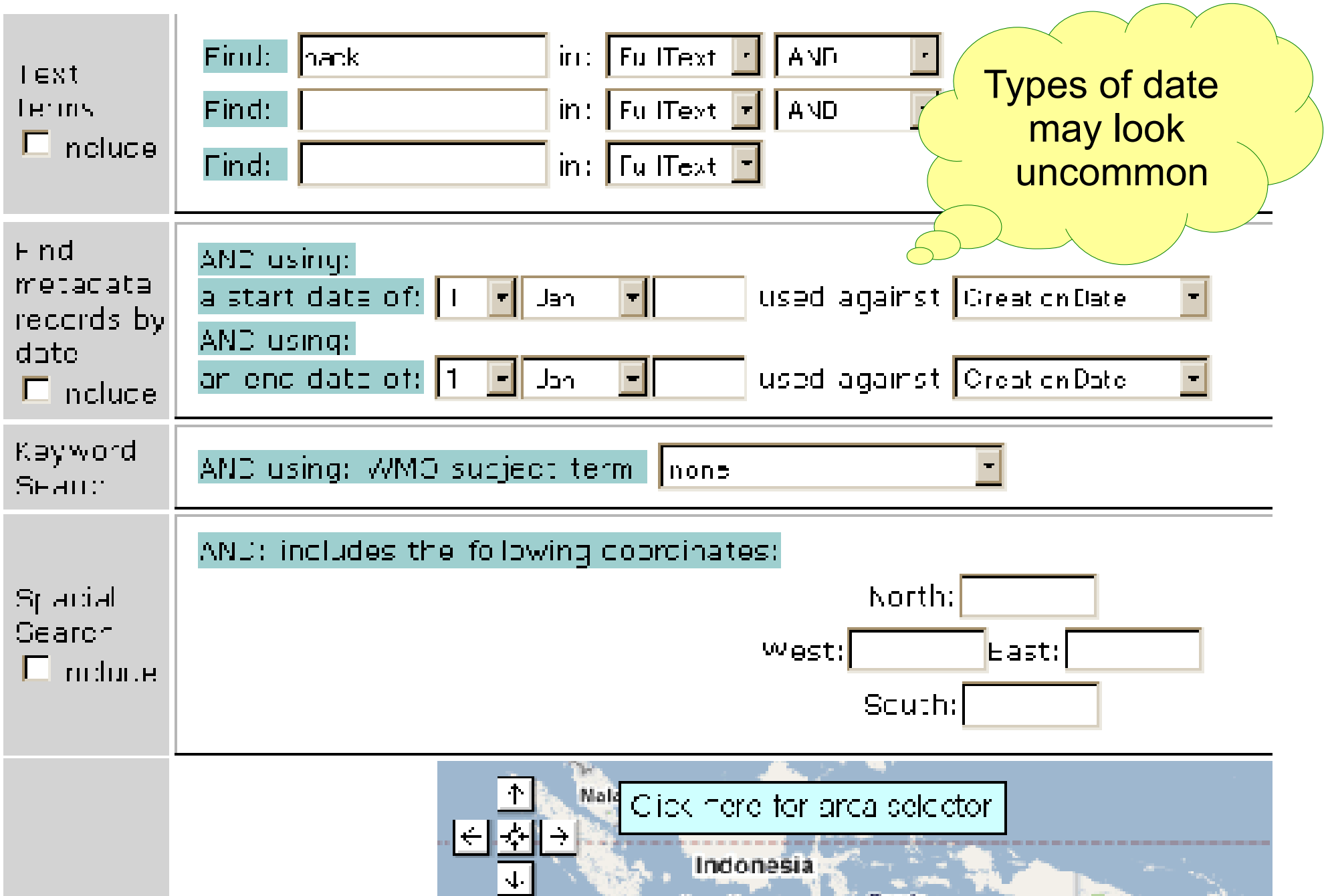
- author
- title
- abstract
- identifier
- keyword
- type
- crs

- TIME
- creationDate
- modificationDate
- publicationDate
- beginningDate
- endingDate

LOCATION

- bounds

Search form at WMO web site



Inter-organisational metadata exchange

OAI-PMH protocol

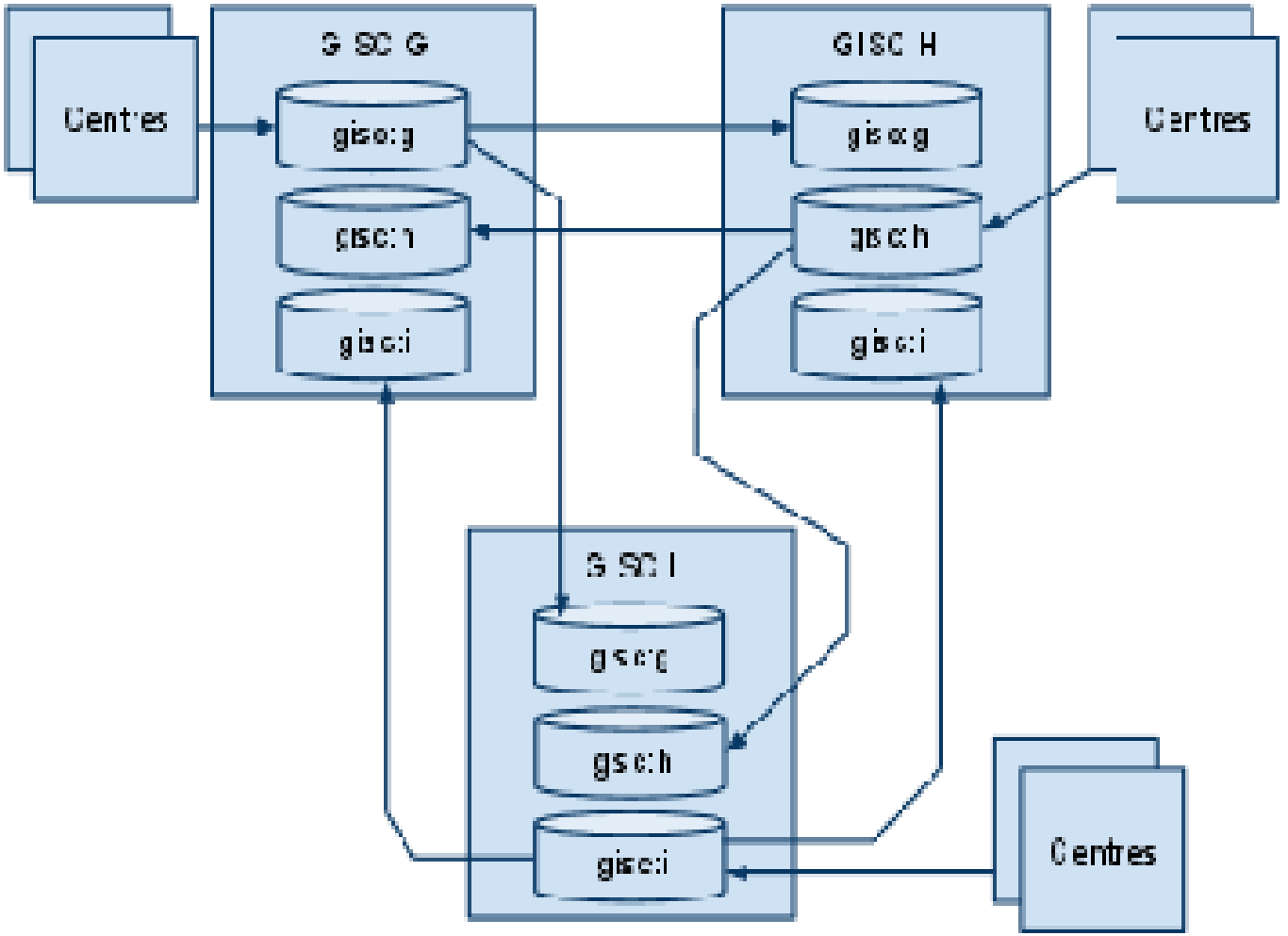
- Developed in library community
  - Update notice like in RSS
  - Time range specification (to avoid overflow)
- Technical features similar to SRU
  - Request: CGI-like parameters
  - Response: XML including metadata records

OAI-PMH in cyclic topology

- GISCs harvests each other's metadata
- Originally OAI-PMH is designed for non-cyclic network
  - An identifier might be inserted by any GISC
  - Whichever looks new copies to another
  - Copied record must have further newer timestamp
    - That is mandatory to avoid missing record
- Hence any update loops!

Convention to avoid cyclic harvesting

- Records divided in sets, only one harvested



OAI-PMH interoperability test in WIS

- Many implementations
  - jOAI-based ones
  - Geonetwork
  - Scratch-up
- Monitoring, finding & solving problems
  - Compatiblity of GML namespace
  - status="deleted" not supported by Geonetwork

GML namespace issue

- Geographical metadata standard ISO 19139 uses GML (ISO 19136)
- Text refers to ISO 19136
  - xmlns:gml="http://www.opengis.net/gml/3.2"
- But XML schema (published in 2007) had used draft ISO 19136
  - xmlns:gml="http://www.opengis.net/gml"
- Former one agreed standard
- JMA provides online conversion service

Next step

Efforts to maintain catalogue

- Many systems left unupdated
- Existing catalogue activities must be
  - Use
  - Take over
  - Improve

GTS bulletin catalogue

- CSV format
  - 12 columns: some have obscure meaning
- Use: now metadata is made from
- Take over: someday
- Improve: table made from metadata could have clearer definition
  - Many properties could be extracted from self-describing file formats
  - Manual input needed only for title, abst, use limits

Thank you

Pre-operational at www.gisc.kishou.go.jp