



Meeting Notes and Actions

Revision history
Draft 0.1 (BNL, June 16)

These are notes describing issues and actions from the breakout sessions of the GO-ESSP meeting at BADC in June 2005.

CF Breakouts

1. CF Broad Issues
 - a. Technical
 - i. Mailing List, Audit Trail Mechanism (CVS?), Website, “Standards-Forge”, Bugzilla
 - b. Social, need:
 - i. “Guardians”, active “Secretaries” (CF itself, standard names ...)
 - ii. Due process, both internally and wrt other organisations
 - iii. Decisions on importing versus copying, and componentisation
2. CF Issues
 - a. Staggered Grids
 - b. GIS Info
 - c. Ontologies and Semantics
 - d. Discovery Info (a la Ethan)
 - e. Coordinate System Object Model
 - f. Scientific Data Types
 - g. Station Data
 - h. “Measured Parameters” (e.g. radar signal strength, which have yet to be turned into geophysical measurements).
 - i. Dictionaries (a la Roy).
 - j. What does CF compliance mean?
3. CF Funding
 - a. BADC hope 50% FTE from 2006.
 - b. Subscription Mechanisms? The Consortium Approach. Split organisation management concepts from science management concepts.
 - i. Org Management:
 1. Who holds the cash?
 2. Voting? Regional? Geographical Balance? Community Balance? Discipline balance (CF dominated by modellers at the moment, need more obs input?).
 - c. Benevolent Organisations (contributing in kind):
 - i. NCAS (BADC – tbd?, and CF checker deployment, CGAM, CF checker maintenance)
 - ii. Unidata (reference implementations?)
 - d. Can we get “Favoured Nation” status from WMO?
 - i. Feeding in and receiving blessing from WMO
 - ii. Get to the Commission for Climatology
4. CF Feedback from Group A
 - a. Yes: to separating out standard name management from the rest
 - b. Yes: this is a GO-ESSP issue (at least at this meeting where we have funded the CF authors to attend)
 - c. Yes to modification methodology:
 - i. Proposal Volunteered
 - ii. Discussion

- iii. Provisional Resolution
- iv. Trial Implementations
- v. Re-assess
- vi. Harmonisation, Public Review
- vii. Decisions (relying on consensus)
- d. Implementations are key to this. Test applications the most complete appear to be the IDV and CDAT. Others include Ferret, Ingrid, general java clients.
 - i. Need a reference library.
- e. Should backwards compatibility be sacrosanct?
- f. Module Compliance rather than entire compliance?
- g. Justifications for compliance: Application building, Curation
- h. Priorities (for funding deployment)
 - i. Moderating the mailing list (taking over from Brian)
 - ii. Issue Tracker (minimally just a threaded email site plus simple issues list)
 - iii. Name maintenance
 - iv. Web Site
 - v. Technical Writing
 - vi. Reference Implementations (can we prevail on unidata)
 - vii. CF checker (can CGAM and BADC continue to maintain/deploy these?)
- i. Guardian Committee should include at least: modeller, data manager, software developer, observational expert, chair. Note that no casting vote should be necessary, "No consensus, no decision".
- j. Solutions must bridge the Atlantic.
- 5. CF Feedback from Group B (additional to that above)
 - a. CF version 1.0 needs to commit to data that has already been written
 - b. WMO/IPCC authority and WGCM committees need to be aware and involved.
 - c. Speed of development should not be reduced by process
 - d. Annual GO-ESSP could include a CF workshop component.
- 6. CF Feedback from Group C (additional to that above)
 - a. Process needs formalisation
 - b. Need to identify clearly the communities and ownership (biology & chemistry, in or out?, if out, how to include?)
 - i. impacts community?
 - ii. modellers (global, mesoscale, finer)
 - iii. observational
 - c. Core and special sections even within vocab? Exploit namespaces better.
 - d. Directory of CF tools?
- 7. CF Decisions
 - a. We say the current version is 1.0. However, station and trajectory data support should be unwrapped.
 - i. check this with the mailing list.
- 8. CF Action Timetable
 - a. White Paper (first draft, end of July, Bryan and Steve)
 - b. CF authors to review (August)
 - c. Public Discussion (September)
 - d. CF authors agree to hand over to new structures and timetable
 - e. Establish New Structures (committee memberships, October)
 - f. New Committee takes forward consortium proposal
 - g. Lightweight Issue Tracker (November)
 - h. Onto CF 1.0 Review ...

Meanwhile, everyone to try and get institutional buy in.

Discovery Breakouts

9. Notes on WMO Activities (from Gil's ad hoc explanations):
 - a. WMO core schema 0.2 is an implementation of ISO19115 (itself just a content standard).
 - b. Done without GML
 - c. There is an ad hoc group (who? Gil to publicise) running this.
 - d. Feed experiences and proposed changes to Gil/Luca/Jeremy
 - e. Note that SIMDAT not using it internally, using "Travelling Metadata", but outsiders will see ISO19115. (TM is to reduce volume for GTS type data granules).
 - f. Concentrating on real time data.
 - g. How does it do the link back to the data?
 - i. This is the key implementation issue a la DIF/Thredds etc.
 - h. Note Beijing meeting in September.
 - i. Gil to find out how open this is.
10. Linking metadata to data.
 - a. Specific data URL addition to DIF has been proposed.
 - i. should this actually be a collection of services?
 - ii. should it have both a service type and dataset identifier, not just a URL?
 - iii. Difficult to keep up to date?
 - b. Two paradigms to consider: real time v static.
 - c. Need Global Identifiers
 - i. need to be hierarchically managed, authority to grant authority, requires resolvers?
 - ii. barcodes?
 - iii. have been pilot projects ad nauseum, can we just do something?
 - iv. Z3980 register unique identifier systems
 - v. DOI ?
 - vi. Gil to fish out a WMO proposal in this area?
 - vii. Need a clear idea of the domain over which the identifier is expected to be unique.
 - d. Issue of Aggregation dataset identifiers and how one handles them in discovery.
 - e. Publication and Provenance need to be included (Lineage, and lists of constituent identifiers).
 - f. Luca to make some proposals (based on some principles that we can agree, and then exchange records based on).
11. Proliferation of data copies. How does a search engine cope?
 - a. Order response by provenance (not linear)
 - b. Publish a dataset, and others should expose copies, and distinguish between.
 - c. Data copied from one place to another, then the identifier should remain the same, what about if the serialisation differs (well, if the dataset is no longer bit-compatible), then it's not the same one. But heritage belongs in provenance.
 - d. Hashcodes /MD5sums etc/Payload
12. OAI
 - a. We should aim for automatic OAI systems by next meeting. NCAR, NDG plus whoever else.
 - b. Avoid replication.
 - i. Unless original unobtainable (firewall)
 - ii. OAI has a friends indicator (identifier, need to exploit this)
 - iii. Registries
 - c. Discovery needs to consider layers, drilling, browsing
13. IOC Relationship
 - a. Note Greg Reed has been charged with developing a marine profile of ISO19115 (in the MEDI context?)
 - b. Two meeting in the UK in late July (vocab issues @ BODC, structure issues in Exeter).

Model Metadata

14. Numerical DIF extension
 - a. Needs circulation. This will happen as soon as Bryan writes it up.
 - b. Generic Issue: how to extend DIF locally, what is the best practice?
 - i. Use of container schemas? Michael to thresh this out in a longer email.
15. Model metadata important (here aimed towards consumers rather than reruns or for couplers).
16. IPCC testbed of concepts.
 - a. Earley Suite
 - i. Start by promulgating "V1" of the Earley Suite Schema (including some Balaji concepts)
 1. improve the time elements
 - ii. Apply this by hand to the IPCC datasets
 - iii. Produce xml and a db interface and hand back to PCMDI as a showcase for the new schema.
 - iv. Aim to do this by the end of summer.
 - b. DIF extension (aka Summary Model Metadata)
 - i. propose to GCMD in parallel with
 - ii. test by summarising the results of the Earley Suite IPCC work
 - iii. try and produce an XSLT from the Earley Suite docs to this.
 1. (NDG will try to do this)

General

17. Original "working groups" (Access Control, IO Subsystems, Model Metadata, OAI Harvesting, Search and Discovery, Standardisation) no longer quite ideal. Better to think about the "pilot projects".
18. 2005/06 GO-ESSP Pilots (and named "leaders" include):
 - a. access control (NCAR/NDG experiments to take place in early 2006, Luca/Phil)
 - b. metadata
 - i. Simulation Summary (for DIF) (Jamie)
 - ii. Simulation Metadata (Lois)
 - iii. CF colocation (tbd: chair of the CF committee)
 - c. discovery (Michael)
 - i. OAI harvesting best practice doc (namespace issues, tutorials)
 - ii. OAI harvesting "in place"
 - d. standardisation
 - i. WCS/WMS experiments (watch Galeon, encourage Unidata to come to next GO-ESSP to talk about this)
 - ii. all working on this
 - e. portals
 - i. Live-Access Server Developments
 1. security
 2. restructuring.

Next Steps (apart from pilot projects):

- Bryan to circulate email addresses of all attendees (underway)
- Bryan to collect all presentations and provide to Chris for linking to agenda at website (underway).
- Gil to provide more info on WMO core group and mechanisms for getting involved, also to find a WMO doc on unique identifiers.