SPG meetings
University of New Hampshire

Tuesday July 15, 2008- SPG meeting

Attendees: Rich Ullman, Yonsook Enloe, Allan Doyle, Helen Conover, Glenn Cunningham, Louis Kouvaris, Karl Benedict, Ed Armstrong, Gao Chen, Mary Kleb, Mike Brewer, Brian Sauer, Karen Zanter, John Scialdone, Siri Jodha Khalsa, Anne Ball, S. Daniel Jacob, Rob Raskin, Deborah Smith, Hongliant Fang, Raj Singh, Ted Habermann, Ken McDonald, Frank Lindsay, Howard Burrows, Ken Casey

Rich Ullman
Gave an introductory talk on the SPG background, the standards process, and the SPG experiences;
Q: Any standards that have gone nowhere?
A: GeoTIFF – could have used funding to provide support for writing of this.
Q: NetCDF vs. HDF – contradiction to recommend both?
A: Not exclusive recommendation. Also these two are merging.

DISCOVER (www.discover-earth.org)- (Deborah Smith)
Provides accurate long term climate data; processing techniques; online services for data access and visualization; merged products: OISST, Water Vapor; using binary format; GHRC data pool (common user interface) at datapool.nsstc.nasa.gov; integration of OGC web services; Web and FTP distribution
Q: What standards to pick up or share?
A: Using some already approved. May be some interest in writing up binary format. Better if there was some attached metadata (e.g., header).
Q: How is read program different from format library?
A: Much smaller.
Q: What about long term maintenance?
A: Documented on the web site. Data to go to NSIDC long term.

SWEET (Rob Raskin)
GCMD science keywords vs CF Parameter Names; More scalable/more descriptive/multi-faceted; ontology representations: Subject-verb-object
- SWEET 2.0 moved from 12 "boxes" or facets to ~80
- **Want to submit as community standard: particular version + procedure for updates.**
  o ESIP Federation Semantic Web cluster to manage and maintain
  o Looking at rules for creating discipline ontologies – but how to enforce?

Multi-mission integration for ocean altimeter data (Daniel Jacob)
Q: Why NetCDF?
A: Small files, so HDF may be overkill. NetCDF enables addition of metadata to binary data – long term solution.

Time-enabled WMS in Google Earth (Karl Benedict)
- WMS 1.1 included time as an optional capability. KML (Keyhole Markup Language) 2.2 is the OGC standard. Images representing geospatial data (gif, PNG, JPEG); temporal requests; GetMap request (format, size, multiple layers, transparency); sample time-enabled layer block from the capabilities XML file; add a time request (point in time or a span)
- GE, udig, arcGIS all support GetCapabilities response display

- Web based mapping client – RGIS data viewer
- GRASS raster database for imagery

LCDM Data Processing and Archive Segment (Karen Zanter and Brian Sauer)
LDCM July 2011; operational land imager (9 spectral bands); thermal may be included; USGS HQ policy – L1T product available to everyone after registration; full resolution browse, quality band and metadata available through WMS; use WCS for single scene data delivery;
- **Big GeoTIFF** and WCS for L1 data, **GeoTIFF** and WMS for browse
- HDF-EOS5 for L0 archive (considering compression)
- Still researching metadata standards (GML or a subset, geoRSS, KML, CSW)
- Earth Explorer is default data access tool
- SOAP / WSDL / XML, but not using UDDI (no one is). Keeping a list of services (name + WSDL URL) – manually generated now which helps enforce versioning. Could harvest WSDLs to generate service list automatically.
- Standard products vs. generation on the fly. What framework to use?
  - OGC looking at BPEL and SWE for this

Discussion of what to standardize next
- **GeoTIFF** (Landsat and ASTER using this) – but what to standardize (GDL implementation is most widely used)
- What about a NASA profile of FGDC (content) or GML (structure) for metadata

OCEANS SESSION

Jason 2 mission standards (Ken Casey)
OSTM: standards through negotiation; using the OAIS for archive; Submission Agreeemtn with all the partners – have something better than before. Agree on standards; the OAIS Submission Agreement is very useful for specifying the process and the standards that will be followed; The submission agreement took 3 years to negotiate and specify although the actual amount of work done was relatively small. Led to more robust, standardized, useable, and archive-ready data. The schedule for drafting of the Submission agreement was dictated by the mission launch schedule. Starting the agreement process early allowed for sufficient time to reach broad, international consensus and a much more robust data stream.
- Using various data and metadata standards (FGDC, NetCDF, CF, standard names)
Q: Standards to make archives better and more efficient, or to help users?
A: NetCDF / CF addresses user and archive needs. Services on top to meet specific user requirements. (Not just preserving data, but preserving for a user community.)
Q: what about preservation planning?
A: OAIS pushes you to consider this. Note that data heterogeneity (i.e., data format controlled by data provider) is a long term issue to consider in preservation. Archive assumes sufficient control of the information to insure long term preservation.

Group for High Res SST (http://www.ghrsst-pp.org) - (Ed Armstrong)
International project creating merged SST products (18) – incomplete to complete data; level 3 data (gridded L2P); Level 4 data – gridded and combined; offer 12 TB of online data with 5 million files

downloaded last year). Ed has been involved since 2002. GHRSST project – partners, products, distribution; In principle, the merging and analysis of complementary satellite and insitu measurements can deliver SST products with enhanced accuracy, spatial and temporal coverage. 35 GB and 6500 files transferred from GHRSST data providers for each data. Currently 12 TB per year total rate. 3.5 TB of products are available for download from the po.daac ftp site.

Product specs – 2003 – using netCDF 3.6 bundled with CF/COARDS compliant metadata. Have not seen any netCDF files that didn't have at least some CF being used. Metadata specs (2003) – DIF plus some custom tags. A granule level xml metadata for each file made by producer. Using DIF tags for ganule level metadata.

- NetCDF 3.6 with CF/COARDS (advantages: direct access, etc.; disadvantages: no internal compression, externally compressed, no chunking
    - o Looking at NetCDF 4 – benefits: internal compression and chunking, support for parallel i/o, backward compatibility with v3.6. NetCDF 4 will be evaluated this summer with a decision by the next GHRSST science team meeting in June 2009 for adoption.
- DIF compliant metadata ; granule level XML metadata (file record for each file); product level XML metadata; some "custom" tags added; JPL Master Metadata directory;
    - o File records and dataset descriptions follow DIF XML schema (with some custom tags)
    - o Convert to FGDC via XSLT style sheets. Creating file level metadata for FGDC too, with multiple file records in same FGDC record
    - o Looking at ISO 19115-2 (probably mapping from current schema)

Future metadata requirement – ESA will require metadata for next generation sentinel satellite derived products to be ISO 19115-2 compliant. This directly impacts GHRSST.

DMAC (Ann Ball)

DMC goal: infrastructure (data discovery, access, transfer, metadata, archive); DMAC is one of three IOOS subsystems; DMAC Steering team consists of expert teams, caucuses (international, private sector, education, modeling, regional); interagency coordination; Teams: review standards, make recs, identify gaps; organizations: implement standards and recs, fill in gaps; Standards : identify standards needed, coordinate with other standards processes, adopt, adapt, and develop; DMAC stds process- 4 moved to "proposed"; need better instruction and/or help for submitters

- Expert teams and steering team are advisory only, no implementation at this point. (Note systems engineering working group is not active.)
- "standardize a standard" – i.e., develop a profile

Q: how to map metadata across different groups
A: mega-metadata-martix – looking for more help with this
Q: Ted points out that this is a new standards process, but DMAC has been evolving over several years
Q: Why to submit to this process?
A: Get comments, advertise through community, may be included in grant language
A: Probably won't spend time on federal mandated standards (FGDC, DIF, ISO 19115)
Q: Coordination with NASA process
A: may fast-track NASA recommended standards (don't bother with tech reviews)
A: Note that NetCDF (currently in SPG process) is not going through DMAC alone at this point, but in combination with CF, OPeNDAP and aggregation
Q: Any push-back on packaging standards for review?
A: Strong opinions on both sides

NASA HQ Endorsement Process – what happens next? (Frank Lindsay)

- Ken Casey – wouldn't an ESIP Federation SPG be better, with NASA, NOAA, others all subscribing?
  - NASA has a unique mandate (science missions) that's taken very seriously. Standards have budgetary impact; should provide risk reduction / cost savings. Might influence peer review process.
  - NASA helps fund Federation and has an interest in activities
  - NOAA tends to respond better to external requirements and standards, might appreciate a community standards process.
- Yonsook – how would that be different from ISO, FGDC, OGC?
  - NASA involved with all of these
  - These are standards development orgs (more top down)
- Ted – many groups developing standards processes, but not really approving standards (maybe 1 per year).
  - Elena said submitting HDF5 was hardest thing she'd ever done. (Note James had positive response to DAP submission.)
  - IETF model is maybe not the best model.
    - IETF members are not volunteers. (Neither are SPG members – but there's no way to fund RFC developers or reviewers).
    - Other reasons that Ted can't remember
  - Ted thinks that clearly defined metrics might help us evaluate how well our process is working.
- **Maybe stronger endorsement from NASA HQ would be better incentive to submit standards**
- Looking for stable of standards applicable to NASA data (not just one).
- Good standards propagate through community. E.g., international airborne data community wants to share data, so these investigators are encouraged to make data interoperable and registered.
- **Data producers / systems should have to use recommended standards or convince peer reviewers why those don't fit. Then what?**
  - **Do we solicit RFC describing the new technology?**
  - Maybe getting around "recommended standards" will be too easy
- SJSK points out that no standards have yet been approved by HQ
- Ted – if process is too long, standards will be superseded before approved
- SJSK – requirements process – what kinds of standards does NASA need?
- Marginal costs to picking a standard (or changing to a new one).
  - A driver is that global climate change requires merged data – compatible and interoperable. Incentive to use standards is that data can be more widely discovered and used.
- How many of standards out of this committee have actually been approved / recommended by HQ?
  - Frank doesn't know – new on the job
  - Who would recommend standards – CIO? Maybe preferable to keep within ES Division of Science Mission Directorate
- Focus on NASA so we can improve relationships with other agencies

GEO Standards and Interoperability Forum (SJSK)

- "Interoperability arrangements" instead of standards. "Special arrangements" if protocols used are not already on the list of interoperability arrangements. Special arrangements may be promoted.
- Hope to foster convergence on limited set of standards. Maybe publish number of implementations or instances of each.
  - o Allan: this list will never converge, only explode
  - o Frank: difficult international context – maybe can filter and organize standards list in some meaningful way

Discussion of what to standardize next
- **CF.** Maybe Russ Rew will submit after NetCDF goes through. Currently managed by UK Met Office.
- **KML**
- **Temporal WMS** as potential tech note – Karl Benedict

July 16 KML/GeoRSS lunchtime BOF
Approximately 30 people attended. Many signed the attendance register:
Allan Doyle, Deborah Smith, Brian Wilson, Karen Moe, Steve Olding, James Gallagher, Steve Young, Erin Robinson, Yonsook Enloe, John Scialdone, Clyde Brown, Liping Di, Rich Ullman, Raj Singh, Pat Cappelaere, Karl Benedict, Siri Jodha Khalsa, S. Daniel Jacob, Ed Armstrong, Helen Conover, Saurabh Channan , Tyler Stevens, Rudy Husar, Wolfang Grunberg, Danny Hardin

More than half use KML or GeoRSS

KML Users: Remote Sensing Systems (for display of SST data)
KML is used to browse data and metadata
GCMD: DIF to KML converter
Brian (JPL)- Formalize KML use through Federation?

GeoRSS: Enable searching mechanisms. Add metadata to the RSS feed

OGC: WFS to KML
        Style Layer Descriptor rules
        Bi-directional between GeoRSS and KML

KML/ GeoRSS Splinter meeting joing with The Infusion organized during the October meeting.

Karl will work on Lessons Learned/ Best Practices Paper

Brian (JPL) and Karl: ESIP lead microformat activity.

In October, plan to have a GeoRSS/KML session – maybe a joint session with tech infusion – maybe half day of meetings.