July 17, 2007
SPG meeting at Madison, Wisconsin
9am

Attendees: Rich Ullman, Yonsook Enloe, Ed Armstrong, Allan Doyle, Tyler Stevens, Jingli Yang, Dave Danko, Ananth Rao, John Evans, Peter Vretanos, Cynthia Wong, Ted Habermann, Ken Casey, Manil Maskey, Helen conover, John Scialdone, Daniel Jacob, Brandon Badger, Ethan Davis, Marilyn Kaminsky, Stephen Bernick, Jeff Arnfield

Attended for short time: Liping, Meixia, Peter Fox, Howard Burrows, Danny Hardin, Kathy Fontaine, Mary Wallace

Notes from Discussion:

Google Map API (Brandon Badger – Google)

- HDF5 being adopted by Matlab as their base file format
- NetCDF4 Java re-implementing a part of HDF5.
- Google Map is all about locating and displaying spatial data. All KML files are indexed by Google web crawlers. KML now readable directly by Google Maps API. Google moving torwards making KML an open standard so there will be more data indexable for GM and GE.
- Sketchup (3d modeling/sketching modeling tool), google maps for mobile devices
- Google Earth – people can overlay your own info (photos, text, kml layer) onto google earth – other people can then search on it and access your info layer; KML files can link to each other to represent arbitrarily large data set; easy to share KML files – users benefit fronm common navigation UI in google earth; level of detail, zooming in/out; KML great story telling tool; KML can get data into public consciousness – to make data matter to the public – not just in scientific circles; Google earth is a storytelling tool;
- Google Maps is the way to present Geo content in a web browser – not 3D; Javascript API to facilitate mashups; google maps & craigslist housing & create something even more useful; ideally the new site adds value;
- Until now - - KML for Earth & Maps API for Google Maps based hmtl pages;
- KML and GeoRSS layers support for the Maps API – then if you use KML, still can get your data into Google Map; KML- Network link example – can do dynamic layers that way –set time to live parameter to refresh: KML is the Common Geo language; KML support for google maps;
- KML 2.0 – specify icons and labels to identify locations on the planet surgace; create different camera positions; …..textured 3D models ; incremental updates;
- KML – started as keyhole markup language – google is giving up control of KML. Is working with OGC to make KML a best practice. Is conducive for indexing geo data if more data in KML. KML 3.0 will be OGC spec. KML good for displaying data but not a data format itself. Can take WMS layers and display in KML.

- Google maps support GeoRSS. There is some overlap between georss and kml. Room for multiple standards; trying to create tools that will help people create their own geospatial layers
- Some experience with satellite data being displayed in KML.

ISO 19115, 19130, and North American Profile (Dave Danko – ESRI)

- Dave Danko – ISO 19115 Project Leader – data about data; producers explain their product and users learn about products; non standard products can be understood.
- Metadata applications: 4 functions: (1) Employ (define, apply, use & understand); (2) Locate (find, discover, structured searches vs unstructed); (3) Evaluate (restritions, quality, reputation) (4) extract (order, download)
- 19115 has topic codes (e.g. ocean, environmnet, etc.) Recommended core 24 metadata elements (half mandatory).
- In 2003 – ISO 19115 --- the schema required for describing fore information and services;
    - Defines metadata elements; provides schema & UML
- 19115 core attributes – contains elements answering the following questions – does a data set on a topic exist: what? for a spec where? Where? For a specific date or period? When? A p0oint fo contact to learn more about or order the dataset? Who?
- Dublin core elements – 15 elements – all optional- loosely semantically defined. Fits a lot of different application. But a lot of different interpretations of what these fields mean; ISO core elements not in Dublin core – mainly to do with spatial info
- What about reference implementation? Not ISO concern.
- Ted Habermann: GeoNetwork tool supports simple schema (for mere mortals) which it can translate to full 19115 schema using style sheet; another tool to translate FGDC to ISO.
- 19139 – XML schema implementation of 19115; defines geographic metadata xml encoding. Uses iso 19118 rule sto produce 19139 xml from iso 19115 uml schema
- Takes advantage of GML and other XML encodings
- Free text field – people often misspell. Codelists prevent the mis-spelling.
- Once it's a draft international std, cannot change it a lot
- 19130 – Image sensor models for geopositioning – deinfes the geolocation information that the data provider must supply in order for the user to be able to find the geographical location of the data using a physical sensor model, true replacement model, or functional fit methodology – New name
- Provides detailed sensor models and associated uncertainties for the following types of passic electro-optical infrared sensor – frame, pushbroom, whiskbroom, a set of components
- Defines the metadata to be distributed with the observation data to enable user determination of geopositions from the observations.

- Camera geometries & coordinate system, platform coordinate systems,
- 19115-2 won't describe sensor stuff – the 130 does. Need both for some satellite imagery data
- North American Profile for 19115 – common US & Canada profile – Natural Resources Canada & Ontario Ministry of Nat Resources, CGSB-CoG; USA – FGDC & ____. Co-located meetings with US & Canada. Selected items from 19115:2003; promotion of optional fields to mandatory; extended codelists; Added best practices – added notes to say what you should do and how to handle it – handled in non-UML way. NAP approval late fall 2007. NCITS L1 will send it out thru FGDC for nationalwide review. Takes from 19115 – Part 1. Doesn't have extensions to completely describe satellite data. Content std and not an encoding standard. Imply that you use 19130 – xml encoding.
- There will be a NAP for Part 2 also, in the future. Maybe revise the current NAP to accommodate Part 2.
- Part 2 applies to level 2/3 satellite data. Raw satellite data needs sensor stuff.
- FGDC Remote sensing metadata std – put ECS metadata into it. Part 2 based on FGDC Remote sensing metadata std.
- Thursday metadata session – discuss what we can do with Part 2.
- Ted – KML so thin, not about data preservation or understanding the depth or the data; need ways to take advantage of structure of data to do detailed analysis

CF Conventions - Ethan Davis (UCAR/Unidata)

- NetCDF is file format for array oriented data; defines structure of data; widely used. ..NetCDF has defined some conventions. CF conventions – designed to promote the processing and sharing of data stored in files created with the netcdf api. Can automatically determine the space-time location of variables CF guiding principles – data should be self-describing. No external tables are needed to interpret the file. Conventions developed only for known issues, easy to use. Redundancy should be minimized, and should be readable by humans.
- CF conventions – determine if variables are comparable – standard names & units; locate in space and time (coordinate variables, projections, etc.); focus on gridded data
- CF spec defined for netCDF-3. But much of CF is format agnostic, eg.g CF-styel metadata has been added to HDF data. The netCDF-4 data model has significatnt changes from the netCDF-3 data model. Won't be sure how netCDF-4 will affect CF until we have more experience with it.
- CF community initially dominated by climate and forecast modelers. Community is broadening; oceanography, atmospheric chemistry, biology; Governance – conventions committee, standard name committee (forked off because std names are such a large and rapidly evolving part) www.cfconventions.org
- Rich: Locate, evaluate, extract, employ – CF oriented toward "employ"
- Ted: CF is more about entities and attributes (more about use than discovery). Not like ESMLK – more semantics oriented. Like ESRI community data models, but ESRI models ream or object oriented.
- Implementations: -

    - CF compliance checker
    - Libcf (support read/write of CF compliant netCDF) (unidata)
    - ArcGIS 9.2 reads CF-netCDF files (esri)
    - netCDF-Java (read, some support for write)
    - CF-netCDF being developed as OGC WCS data format profile.
    - CF-netCDF as TDS netCDF Subset service response
- Leader for SPG endorsement process – large, fairly active community;
- Ted -- CF data model for atmospheric community
- CF – can be used with other file formats – like HDF.
- Note that netCDF4, HDF5 and OPeNDAPO are coming together.
- John Caron the main developer for netCDF4 Java access to HDF5.. Waiting for the full release of HDF5 before releasing netCDF4.

THREDDS (Ethan Davis – UCAR/Unidata)

- THREDDS catalog – xml docs over http; named datasets with access URLs (typed for various protocols) – names human readable; Can link hierarchically to other THREDDS catalogs; Datasets can be hierarchically organized which is good for human browsing; metadata at collection (inheritable) or atomic dataset level.
- Catalog services: validation, subsetting, automatic catalog generation, direct metadata harvesting,
- Typically how many entries in a catalogs? Unidata serves up IDD datasets (radars, etc.)
- Working on aggregating granules into datasets for query services – each implementer decides how to do this. Higher level catalogs reference more detailed ones. Generally used for browsing interactively.
- Implementations of Servers: Thredds Data server, OPeNDAP Hyrax server, GrADS Data Server, IRI/LDEO server (Ingrid), NCAR Community Data Portal
- Implementations – Clients – IDV, MyWorld GIS (Northwestern), netCDF-Java toolsUI, NCAR Community Data portal, Thrdds Data Server
- Installations: NCAR/UCAR, NOAA-NCDC, NGDC, PMEL, COLA, FNMOC, BOM, GMU, IRI/LDEO (Lamont Daugherty)
- Leadership? No governming committee; Unidata pushes THRDDS (Ben Domenico).
- How many entries in a typical THREDDS server?
- Querying capabilities for THREDDS?
- THREDDS aggregation – higher level dataset that describes current and incoming data for the collection. Then even if data streaming in, the higher level dataset will contain metadata about the individual granules.
- THREDDS really designed for a realtime environment
- netCDF4 – merge of netCDF3, HDF5, and OPeNDAP data moels.
- THREDDS specification document -- xml representation of the THREDDS catalog. Required THREDDS metadata: name, dataset id, access mechanism. No geographic info is required.

- Ted Habermann – how to represent aggregation in THREDDS with ISO 19115 aggregation?
  "subsetting" means querying for a subset of the catalog (i.e., "select")
- Yonsook – what aspect of THREDDS would you standardize
  - THREDDS catalog spec – reviewed and implemented by various members of the Unidata community
  - Subset, harvest and other APIs?
- Allan – more for browsing than searching?
  - Initial use case, but not only one
  - No bounding box query support, for example (though user interface can implement this)
- Ted – cool to create series of hierarchical ISO metadata objects from THREDDS catalog (implements aggregation of datasets, not available in FGDC).
- Allan – THREDDS to GeoRSS and KML might be more useful
  - Ted – rather, include KML service reference in THREDDS

OGC CS/W & ebRIM- (Peter Vretanos – Cubewerx/OGC)

- CS/W based on OGC Catalog 2.0.2. CS/W is a protocol binding defined in 07-006r1 (others include Z39.50 and CORBA). No one is using the CORBA binding. CS/W uses http as the distributed computing platform – interaction model is request/response. Asynch requests are also supported. CS/W API is patterned after the web feature service. Query language – supports common query language (CQL); similar to SQL where clause; supports the filter encoding spec for specifying query predicates (with extensions). Profiles may define other languages such as XPath.
- CS/W defines a standard API for creating, updating, deleting and querying catalog records; API can be implemented on top of existing servers as well (Z39.50); service requests may be encoded in xml or as keyword-value pair. The spec support http method post and get and describes how to use SOAP (basically message literal). A specific information model is not defined. It is expected that profiles will be defined to support specific catalog information model. Current APs include FGDC, ISO 19119/19115 and ebRIM. OGC has recently decided to make ebRIM the preferred catalog information model – but more about that in the next presentation.
- The OGC catalog service spec defines a set of common queryables and returnables based on Dublin core. To support cross profile queries. All CS/W implementation must support the core queryables/returnables. Typically referred to as csw:Record.
- The following operations are defined in CS/W
  - GetCapabilities – provides service metadata
  - DescribeRecord: allows clients to get a schema description of the catalog's information model
  - GetDomain: allows clients to discover the runtime value space for API parameters as well as other element within the info model. Get list of valid values back for various attributes.

    - GetRecords: primary method for querying the catalog – supports distributed queyr
    - GetRecordById: convenience request for getting records using their ID.
- CS/W operations: (1) Transaction (primary method for creating, updating, and deleting catalog records (push)); (2) harvet – allows the catalog service to retrieve web-accessible metadata and register it in the catalog. Analogous to transaction but performs a PULL rather than a PUSH and supports periodic re-harvesting of the resource
- The CS/W protocol binding materializes the ogc core queryables and returnables as an xml doc based on Dublin core.
- CS/W KVP encoding – keyword value pair – for lightweight access – can't express all the features of the catalog as a keyword value pair.
- CS/W Xml encoding – allows you to access all of the features of the catalog

- Issues for implementation:
    - No rules for re-harvest (not sure if it should be in the CS/W spec, or the application profile, or part of the standard operating procedure for a community
    - Distributed query is not tested very well – CS/W defines distributed query but not tested very well (what happens if there is a loop in the federation?). server does the search distribution and compiles the results. The search request has a parameter to indicate whether or not to distribute search.
    - With the Info model exposed, interoperability can be a problem in certain cases (e.g. ebRIM Application Profile) – big leaning curve because explosing the info model.
    - Request complexity - because xml is verbose – the search request can be very large and complex. There are 2 types of users – professional and the casual user (mass market). The request complexity is squarely for the professional – the pro GIS domain.
- Catalog RWG working on next version (2.1) – distributed query, multilingual, harvesting improvements, better service metadata via capabilities, SOAP/WSDL/WS standards
- Some things being worked on outside the RWG – "catalog simple" movement similar to the "WFS Simple" idea. Use of GeoRSS for pushing content and notification.
- Peter Vretanos – ebRIM Profile for CS/W –
- ebRIM : e-business Registry Information Model – Defined by OASIS. Used by OGC as the information model for the CS/W-ebRIM profile. Don't use all of the model (like auditing and housekeeping parts). Current version ebRIM v3.0
- Why ebRIM? – need to manage many kinds of artifacts and be able to discover them– dataset descriptions, service offers, coordinate reference systems, units of measure, application schemas, map styles and symbol lib, access control policies, sensor description, ontological descriptions, digital rights, organizations and projects.
- Need to associate artifacts with one another (associations) – the classic found data now want to find service that offer that data and vise-versa. Need to classify

artifacts in many ways – classification schemes. Need to collect artifacts into logical groups (packages); need a common interface for a wide range of metadata management. OGC adapted the ebRIM as the basis for a profile of the catalog 2.0 CSW specification

- Classification with ebRIM –
- ebRIM highlights – a general info model; expressed in XML; can be loaded into an ebRIM Registry; is an ISO std (ISO 15000-3); supports the notion of registry-repository (archive); The ebRIM provides a framework for packages and general info model. The specific package then provides an additional framework for the specific attributes and relationships. The ebRIM is a meta-meta model – an additional level of indirection – impacts performance? But does provide additional flexibility and ability to show complex relationships.
- Highlights – Packages – packages encapsulate all the object types, association types, classification schemes, slots, stored queries, etc. Create/Load an ebRIM Package using XML and WRS interfaces; can support multiple packages at the same time; all packages use common "basic" package; Basically a schema for a catalog instance; can re-use associations and extrinsicObjects across packages; OGC standardization focusing on extension packages for geography
    - EO Products Extension Package for ebRIM (06-131r2)
- Yonsook: cost for this flexibility is additional layer of indirection. Any implementations of large catalogs?
        - Performance issue is implementation issue, not logical issue (CubeWerx has implemented catalog with 4 million records, but bad performance after a million or so, b/c of so many joins required by the model)
- Current ISO (19115 and 19119) spec is being converted to ebRIM package for CSW
- OGC wants to standardize package definition
- Discussion papers that matter:
    - Cs/W ebRIM profile
    - ISO 19115/19119 profile (pre-ebRIM profile) – because guys that wrote this is turning this into an ebRIM package
    - Deep search – CS/W api on top of the ISO model for ebRIM packages
- Other CS/W profiles being updated to use ebRIM (19115/19119, FGDC)
- If 3 groups developing 19115 packages for ebRIM, the 3 packages probably are not interoperable because no one way to construct a package.
- ebRIM has high learning curve – higher complexity. Querying ebRIM requires a sound understanding of the model. Typically multiple joins are required to bring back results.
- In order for the ebRIM to work, need to map from local resources to ebRIM.
- Trick is to hide the ebRIM complexity from the users. Ability to do associations and classifications and to overlay semantics over it are the key strengths of the ebRIM. Can discover ideas now instead of records.
- Why use CSW, not ebRS API?
    - Not a good query API (ad hoc query = <SQL stmt>)

    - Not spatially enabled, and no obvious extensibility for this
    - OGC hoping to harmonize with OASIS ebRIM (b/c some of same spec developers)
- OASIS ebRIM implementation uses relational database
- OGC has done mappings of SWE, M*S into ebRIM (see interoperability reports – look for email from Peter)
- Allan – Aren't you just building RDF on top of all of this stuff? Wouldn't it be better to just start over?
- John – are tools available?
    - Portlets are available to allow you to access ebRIM catalogs

CSW Implementations (John Evans – GIO)

- This talk brought up several more issues with implementing CSW. Gave an overview of the GIO ESG.

GCMD DIF & SERF – (Tyler Stevens-GCMD)

- GCMD DIF – Directory interchange format(8 requireed, 15 reocmmended and 12 optional fields); 19115 was adopted by GCMD in 2004..
- SERF (Service entry resource format) – over 1600 SERFS in May 2007
- Exchange of metadata using XSLT between DIF & FGDC metadata
- Harvest THREDDS metadata into GCMD using OAI/PMH
- Ted – does GCMD DIF have enough attributes so that when the user gets to the data, he understands the data? DIF good for collection level discovery but good for using data? Shouldn't GMCD conform to ISO? DIF has driven a lot of standards but maybe the standards process should adopt its descendants (e.g. ISO). DIF is not really the long term preservation standard. Need more than discovery level metadata, need "use" metadata also. Everyone should adopt ISP and start working on tools to make it more usable.

General Discussion at end of day:

- What are our next steps?
- KML – annotation visually – output of some source data and some styling information that you put together- does not impact catalog discussion at all. OGC has committed to making no changes to KML 2.0. KML 3.0 will be a consensus generated document. SPG should take a look at KML 3.0 when it comes out (Jan 2008). Should be pretty similar to 2.0. KML layered data will attract tons of new users. NASA Ames collaboration with Google earth. Disaster response is an another project. MODIS rapid response.
- ISO 19115 stds are the stds of the future. North American Profile of 19115 – take a look then. Can't use ISO 19115 as is. Each group has to identify which fields

the group will support – all the mandatory plus these optional fields. Part 2 should be ready for final draft Jan 2008 and be ready to implement.

- SPG making recs to HQ for NASA Earth science community – but look at the broader community outside NASA too so that we fit in.
- THREDDS CF – CF will evolve with netCDF4. But don't wait until netCDF4 to come out. CF conventions lowest hanging fruit? (Ed Armstrong) Ignore netCDF4 for now (Ethan) but look at CF for now.
- CSW
    - Allan: not ready for operational use. Maybe recommendation is to put more resources into this, esp ebRIM.
- ECHO
    - Operational, single implementation
    - Publish and query APIs available
    - Inventory only, not catalog level metadata
- KML
    - Widely used, but completely different from other geospatial metadata
    - No required fields, but some common fields defined (author, description, spatial tags)
    - Metadata tag not used by Google, but can be used by communities (include std metadata within KML)
    - Rudimentary support for temporal information
    - Allan: KML is annotation level information. KML search won't really impact existing cataloging efforts
    - Peter: KML could be generated from catalog to display information on Google
    - Ted: alternative is GML
    - Peter: KML3 will be generated by consensus in OGC (may be very different from KML2)
    - Brandon: a way to get data in the hands of users
- ISO
    - Allan: ISO content standard should be desired goal. Agree good to put resources into making this more useful, but not by killing off competition. Maybe it's time to build a roadmap.
    - Ted: encourage people to create ISO, and populate GCMD from that
    - Tyler: looking toward North American Profile