**GeoNIS meeting Day 1**

**GeoNIS definitions**

Stand-alone geospatial datasets…

Discussion on standalone options, what everyone is currently doing… where we may be having issues? Etc.

Jamie

Available as zipped shapefiles

Adam

Available as zipped shape files and e00

Jonathan

EML geospatial, file geodatabase run through Inigo's XSLT

John P.

Available as zipped, e00 as well, EML is there, bounding boxes only. Discussion on standardization of naming conventions for extrapolating name of shapefile to .zip file… making sure that there is consistency amongst all data types

Hope

Zipped shapefiles

Mike

Zipped shapefiles with metadata. Lots of no metadata files. Some with. How far do we need to go with FGDC to be sufficient for NIS… how much is enough?

Aaron

Postgres via Geoserver, zipped and exported from there via DRUPAL site. Somewhat lacking in metadata.

Ken

Little in system, but staging it there. Consistency between compressed and uncompressed versions. Data, documentation, metadata all compressed and packaged together. Really excited about Geoserver. KMZ also.

Ryan

Legacy GIS varying completeness, some EML, some lacking projections. Want to get into Geoserver, postgis, etc.

Theresa

Zipped coverages, e00, name of shapefile is hardwired into metadata.

Kristin

Refused to add to Mike’s stuff.:)

Possibility of using comments in / for zip files to carry metadata, May work but not ideal since we don’t want to have intermediary workings with data, automatic is best. Obviously.

The reason we really need to have our spatial data in EML… if not there, people wont find it when searching against pasta. Since this is the primary direction for PASTA, we have to adopt it.

Best practices are super important here.

Discussion on the diagram, basically what do we need to do in the GeoNIS to make it the most productive entity, not recreating any wheels, and making it stable (BP).

Kens thought, hybrid system… shapefiles-map services is ok, but if something changes there will be an issue. Want to integrate services in the long term.

Porter – where is the master dataset? Is it the zip file at the site level; is it in the map service? Is it accessible from either location or only one? What is the master and how do we deal with that, harvester vs. uploaded manually.

Adam – So, if we have a bunch of separate yet communicating services stored elsewhere, basically we defeat the purpose of the NIS. We need to get this @ LNO

From Mark, what is the constraint in working with data from services vs. from actual files?

Bob Waide wants an LTER Gazetteer – ways to add to it… names linked to coordinates.

**Questions for Mark…**

Integration with Map Services

Review diagram

Harvest interval for Metacat

Can you tell it not to harvest if data hasn’t’ changed? (Large datasets)

Can you have more than one URL listed (one for shapefile, one for web service? As other entity?

Timing of workflow updates?

Storage space for GeoNIS? Right now we have 200gigs

ASM plans? Joint workshop?

Mark? Questions about GIS Services? Questions for us?

Follow up later with LNO, James, JVC, on whether or not they will be contributing data or resources to the GeoNIS. Data, specifically, aerials or other data they are collecting for all sites.

Are we committed to keeping all versions of data, in GeoNIS?

**Chat with Mark….**

Pasta ready – quality EML, passes quality/compatibility checks, make sure that for every entity in the document that you have a dataset that can freely flow into pasta so it can be created/stored/copied into pasta

So, though we have entity types defined as spatial raster or spatial vector, it’s not currently set up for being queried. That said we could move that into a query

Blobs do currently work in the NIS

Timelines –

When is the system going to be available to test on it?

Don’t currently support those spatial entities. No compatibility checks being used / not addressed at all.

Mark has his site review first week of June, mostly off radar until then.

Re: entire dataset being re-downloaded with new EML… way to specify in harvest that dataset hasn’t actually changed. Not yet addressed, reversioning of data hasn’t been looked at thoroughly. Currently any change will be re-harvested. Will look at it in the future. It doesn’t scale. Big issue here.

List more than one URL in EML? Web services vs. data downloads. So… not really. Probably should put this somewhere else. A “distribution URL” within a methods step. We’ll have to look at the EML and see where we should all put it, and build this into best practices.

Re… data checking. A blob is a blob is a blob, so you can’t really use many checks on it. Better to set up specific holders for it.

*ATTENTION*

Is the spatial raster / vector entity valuable at all in this context. Think on whether this entity is of any use and relate it back to EML dev group. Does this serve any actual purpose? / Add any value?

Storage space for NIS somewhat undetermined, TB, but who knows, 54 at present.

Inigo - Instead of exporting FGDC use ERSI style and … convert? ESRI2EML?

ESRI is a profile of FGDC. If you export FGDC capture EVERYTHING. EML requires 3-4 extra fields that aren’t in that. These are, however, in ESRI.

Start storing stuff in our own file structure. Start with doing our stuff directly and worry about it later. Test going from EML to pasta and back to GeoNIS via work through; eventually get rid of first link

Regarding harvesting / data transmission to GeoNIS….

Looking at, for sites with server, allowing a direct connection? For others, doing an ftp/harvest option? Unclear on final thoughts on this…

**REFRESH – things to work on**

Do some testing of EML to PASTA to package storage, generate workflow to bring into GeoNIS

Direct connect to GeoNIS from site, will bypass EML, which isn’t great.

What do we need to convert vs. what doesn’t. For instance, for things that need to be reprojected, does that happen at site or at the network? Need to be aware of this as original EML will reference the original data and projection there in… do we have an additional EML for converted version with new projection?

The blob, the GeoNIS

SDE GeoData

Workflows, multiple

For workflow for people not using ArcCatalog… how does metadata get populated through to final user?

Do we need to rewrite EML again after transformations, or do we use another type of metadata, ESRI?

Research if indexing is available in Arc10 Server side. Also consider non Arc (Geoserver) alternatives.

Discussion on where to, if to, include linkages to collection locations of sites. Matter of resolution, all points? SiteDB vs. in the GeoNIS…? How to address or if to address this.

SiteDB is effectively a giant data table. Use a script to create on the fly updated point layer for all collection locations stored there.

Define each of the components of the components of the GeoNIS…

In terms of long-term completion of this project. Things to consider. Assuming we can document this well enough, the final product will take some dedicated time to complete… buyouts? 2 people @ 2 weeks might do it.

**GeoNIS meeting day 2**

**Best Practices - Brainstorm**

EML – minimize what we can find in arc?

Data packaging – zip, add info / read me, name consistency, additional figs, layers

Attribute definitions – name, type, and number of characters

Metadata entry/editors

Symbology

Coordinate system/projections

Data structures

Web services

Geoprocessing

Data submission to GeoNIS

Workflows

Data structure for SDE databases, vectors rasters, models, etc.

**Best Practices - Data Packaging**

Putting stuff somewhere to be harvested by GeoNIS.

Should be either .zip or .targz,

Should contain native GIS data.

Vector – should be shapefile or KMZ

Raster – GeoTIFF, IMG, E00, ASCII, MRSID

Metadata file -xml

ReadMe File

Package meta.xml

Projection file / projection definition

Symbology layers (opt)

Color Ramps (opt)

*USE EML TO NAME PACKAGE LTER BEST PRACTICES*

Layer name – No spaces

Look into ESRI character limits

HOW DO WE KNOW ITS PART OF NETWORK VS. SITE DATA?

NEED TO DEFINE DIFFERENT TYPES OF INPUT DATA

How to identify LTERMapS data sets

In SiteDB, offer place to tag / select your layers from Metacat by packageID

**Best practices – data submission to GeoNIS**

Prepare package, submit to PASTA, done – john porter

Long run – PASTA will control submission process.

Short term – Metacat, keep zip local on site, reference only

Testing – each site will pre-pack dataset package for test

Interim, Metacat, how do we find data?

Create a query on spatial vector / spatial raster

Possibly add tag / field for GeoNIS compliant 1.1

Workflow

Put in temp

Compatibility checker for GeoNIS data package. Can run by the site and again in GeoNIS

Pass/fail-notify

Load data into GeoNIS Database – Workflow

Populate LTERMapS data page for each site

Via SiteDB or Web Form. Each site defines their stock datasets

**Best practices - Data Structures**

Feature Classes

Site Letter Code … AND Bound (Feature)

Feature name must be unique for the site

Test long name

**Best Practices - Web Services**

Make them available to open source

Don’t embed other web services in your web service! Not best practice

Services offered by topology, thematic, or by site? SITE!

*How do we expose the rest of the services to clients?*

Web Access Page!

Example – bundled services

REST LTERMapS

0 – Sites

1 – Boundary

2 – Ecoregion

**Best Practices – Symbology**

Agree on convention of symbols

For LTERMapS – Symbology is our own, we manage

For Other – Layer file preferred for WMS, WFS, to avoid tacky, nasty symbology…

Color ramps or map for rasters

**Afternoon Stylesheet Demo**

Demo on text file as not to screw up spatial data

If not done before, “customize” ArcCatalog options… “metadata” tab. For style select FGDC… something.

Hit “edit” given all kinds of stuff to edit

From conversion tools, metadata… update metadata, select source, then run FGDC to ESRI

Effectively set up a template from the beginning… when updating metadata for other sets start with aid template and import your template

Keywords, one on a line or CSV? Refer to best practices

For contacts, only use one name per contact, no multiple, name only

If they have multiple roles, create a new contact for them also

**GeoNIS meeting day 3**

**Ken Demo on Geoserver**

Open Source Geoserver software, ESRI for any server background?

<http://www.esri.com/software/arcgis/geoportal/index.html>

<http://gptogc.esri.com/geoportal/catalog/main/home.page>

Malpai Portal example from Ken

<http://spatial-web.nmsu.edu:8080/MalpaiPortal/catalog/main/home.page>

**John Porter Demos**

Data Catalog integration with Google maps… locations of all study sites, or by type, dynamically generated KMZ.

<http://www1.vcrlter.virginia.edu/home1/?q=dataCatalog>

Downloadable Google Earth option as well.

KMZ options for locations, height of line describes quantity of datasets at each location. Time frame option allows for sliding bar which shows how many are in each area at a given point in time.

Geoserver demo… showing off back end of VCRs Geoserver.

Arc to Earth – program can change layer file to SLD, xml style file for Geoserver.

<http://www.arc2earth.com/>

$400.00 for Pro version

Open Layers – JavaScript file that consumes layers as it goes, no server associated with it. Is a client of web services, an API if you will. Plug and play… maybe able to generate dynamically.

**Aaron Demos**

PostGres background… postgres plus postgis

Runs Geoserver on top of this, downloadable from server data.

Arcmap2sld – free

<http://wald.intevation.org/projects/arcmap2sld/>

**Jamie Demos**

Demoing server / services at LNO

<http://Maps.lternet.edu/ArcGIS/rest/services>

Can add services into Arc from above list…Suggested training option for creating and maintaining services stored at LNO. Greeattt!

Write up a proposal for this for next round of trainings.

**John Porter spatial raster and vector queries**

Query on all the spatial data, the spatial vectors and the spatial raster content in EML, PDFs available on GeoNIS page (<http://im.lternet.edu/project/GEONIS>)

Total=695  
Vector=184  
Raster=511  
  
The URLs for generating them are:  
All spatial data:   
[http://vocab.lternet.edu/metacatsearch//querymetacatSpatial.php](http://vocab.lternet.edu/metacatsearch//querymetacatSpatial.php" \t "_blank)  
Vectors:   
[http://vocab.lternet.edu/metacatsearch//querymetacatSpatialVector.php](http://vocab.lternet.edu/metacatsearch//querymetacatSpatialVector.php" \t "_blank)  
Raster:   
[http://vocab.lternet.edu/metacatsearch//querymetacatSpatialRaster.php](http://vocab.lternet.edu/metacatsearch//querymetacatSpatialRaster.php" \t "_blank)  
  
But be advised that since these elements aren't indexed searches can take   
10-20 min to complete.

**Other Thoughts**

Euler’s law – nothing is impossible for the man who doesn’t have to do it himself.

ArcPy. To alleviate memory leaks, run an ArcPy Script as a sub script in python, shutting each down after a run of 1.