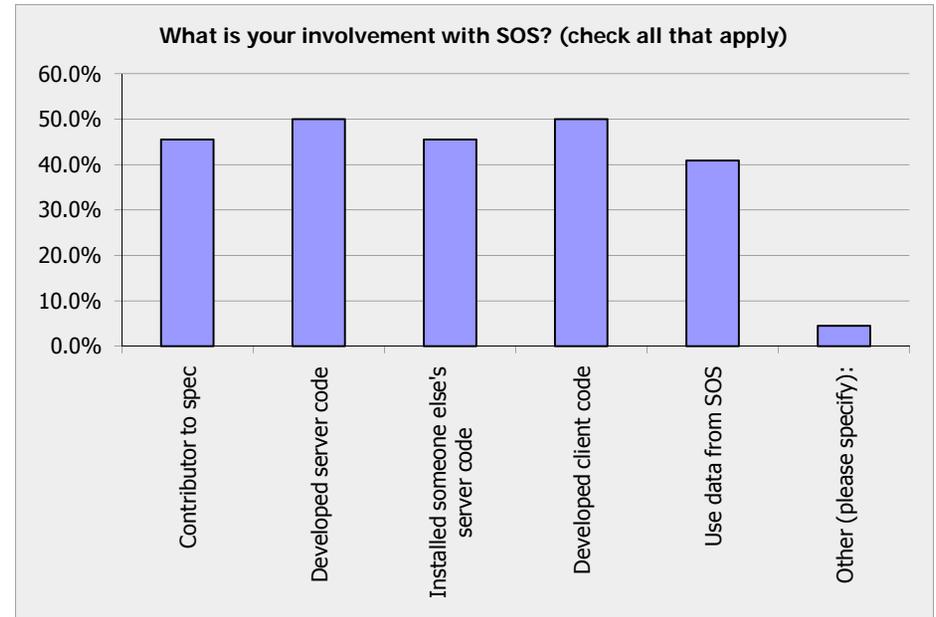


DIF Encoding Evaluation

What is your involvement with SOS? (check all that apply)		
Answer Options	Response Percent	Response Count
Contributor to spec	45.5%	10
Developed server code	50.0%	11
Installed someone else's server code	45.5%	10
Developed client code	50.0%	11
Use data from SOS	40.9%	9
Other (please specify):	4.5%	1
<i>answered question</i>		22
<i>skipped question</i>		1

Number	Response Date	Other (please specify):
1	Jan 15, 2010 5:58 PM	Also enhanced the SOS server code I installed.

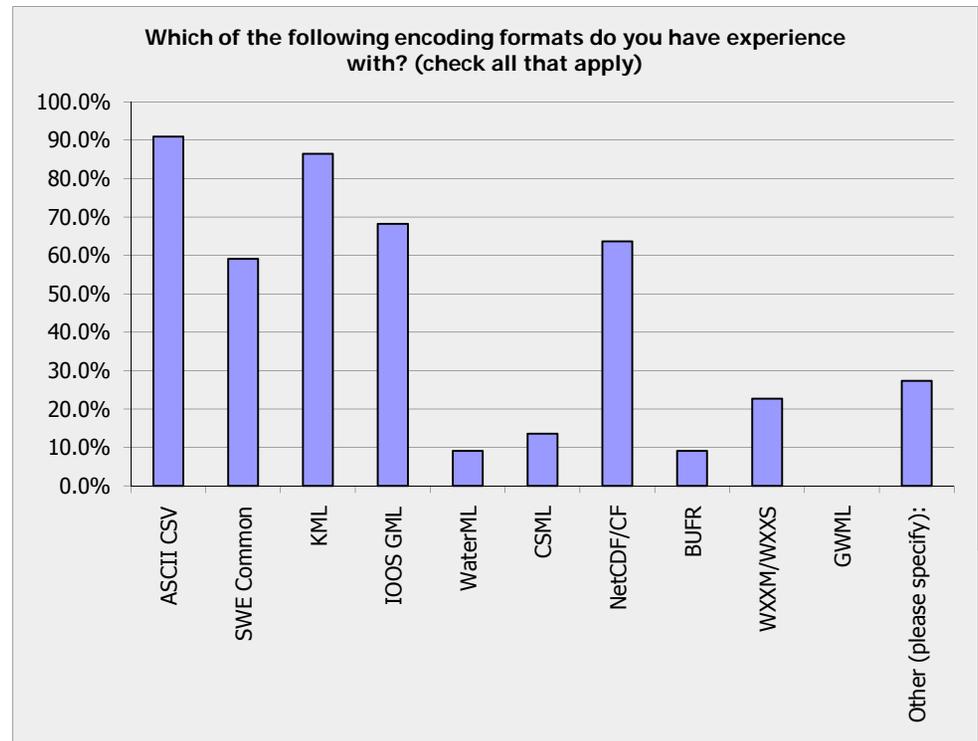


DIF Encoding Evaluation

Which of the following encoding formats do you have experience with? (check all that apply)

Answer Options	Response Percent	Response Count
ASCII CSV	90.9%	20
SWE Common	59.1%	13
KML	86.4%	19
IOOS GML	68.2%	15
WaterML	9.1%	2
CSML	13.6%	3
NetCDF/CF	63.6%	14
BUFR	9.1%	2
WXXM/WXXS	22.7%	5
GWML	0.0%	0
Other (please specify):	27.3%	6
<i>answered question</i>		22
<i>skipped question</i>		1

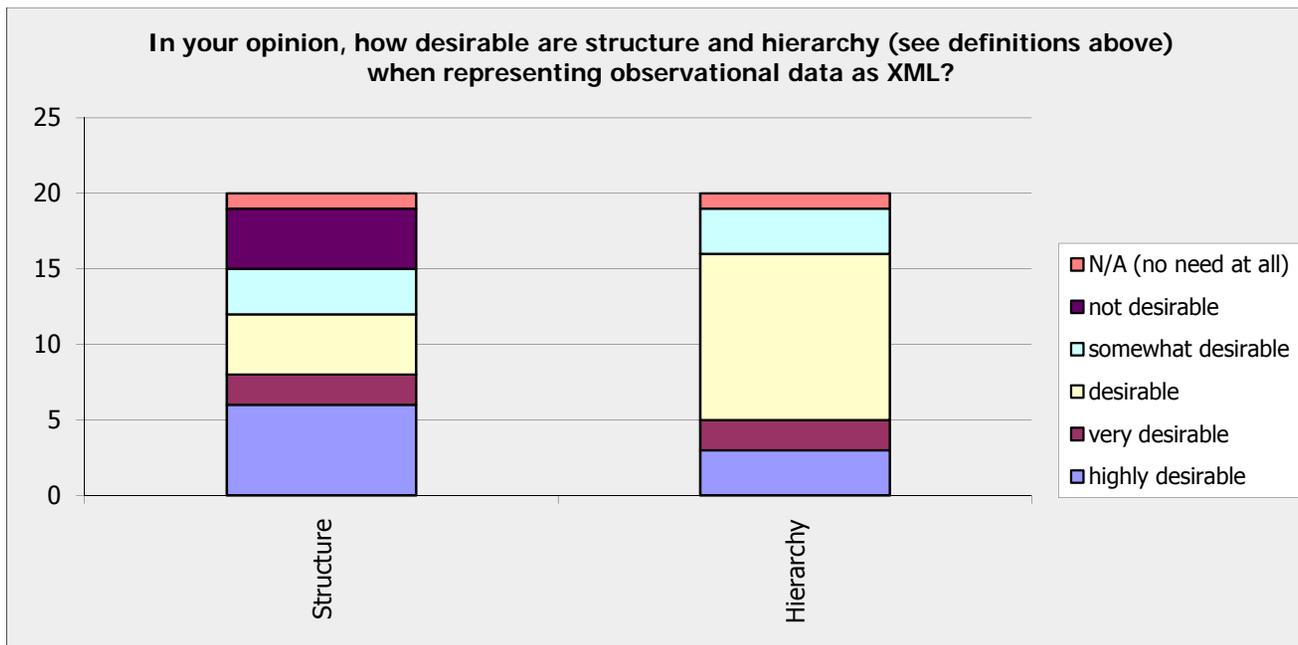
Number	Response Date	Other (please specify):
1	Dec 16, 2009 10:04 PM	many
2	Dec 24, 2009 8:28 PM	GRIB, DiGIR
3	Jan 7, 2010 5:35 PM	Various XML
4	Jan 7, 2010 6:16 PM	HDF4/5, grib/grib2, WMO coded observations and data on the GTS/NOAAPORT
5	Jan 13, 2010 7:40 PM	GRIB and GRIB2
6	Jan 13, 2010 8:17 PM	CDM (Common Data Model)



DIF Encoding Evaluation

In your opinion, how desirable are structure and hierarchy (see definitions above) when representing observational data as XML?

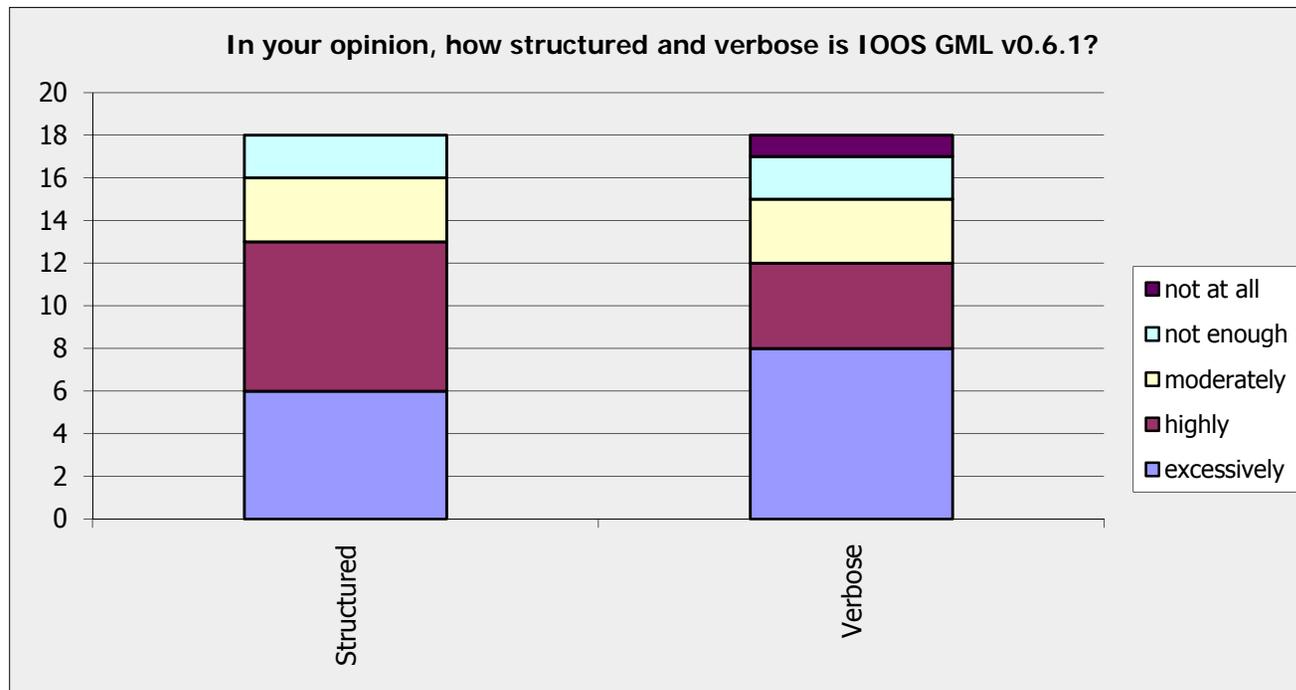
Answer Options	N/A (no need at all)	not desirable	somewhat desirable	desirable	very desirable	highly desirable	Response Count
Structure	1	4	3	4	2	6	20
Hierarchy	1	0	3	11	2	3	20
<i>answered question</i>							21
<i>skipped question</i>							2



DIF Encoding Evaluation

In your opinion, how structured and verbose is IOOS GML v0.6.1?

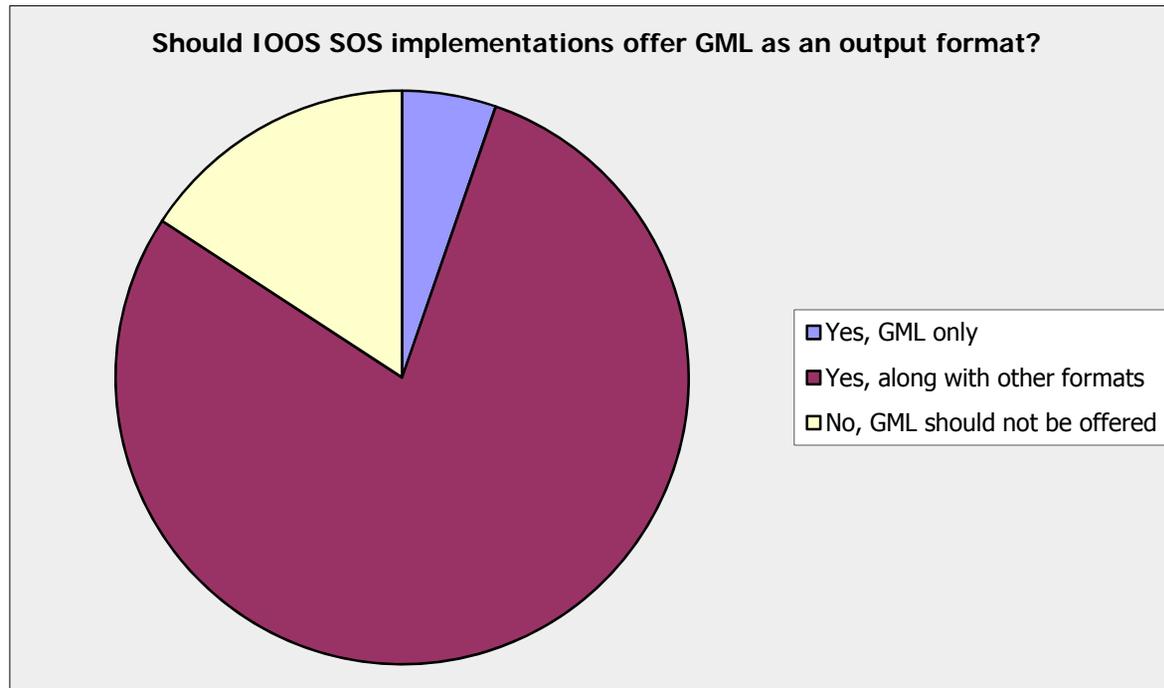
Answer Options	not at all	not enough	moderately	highly	excessively	Response Count
Structured	0	2	3	7	6	18
Verbose	1	2	3	4	8	18
<i>answered question</i>						19
<i>skipped question</i>						4



DIF Encoding Evaluation

Should IOOS SOS implementations offer GML as an output format?

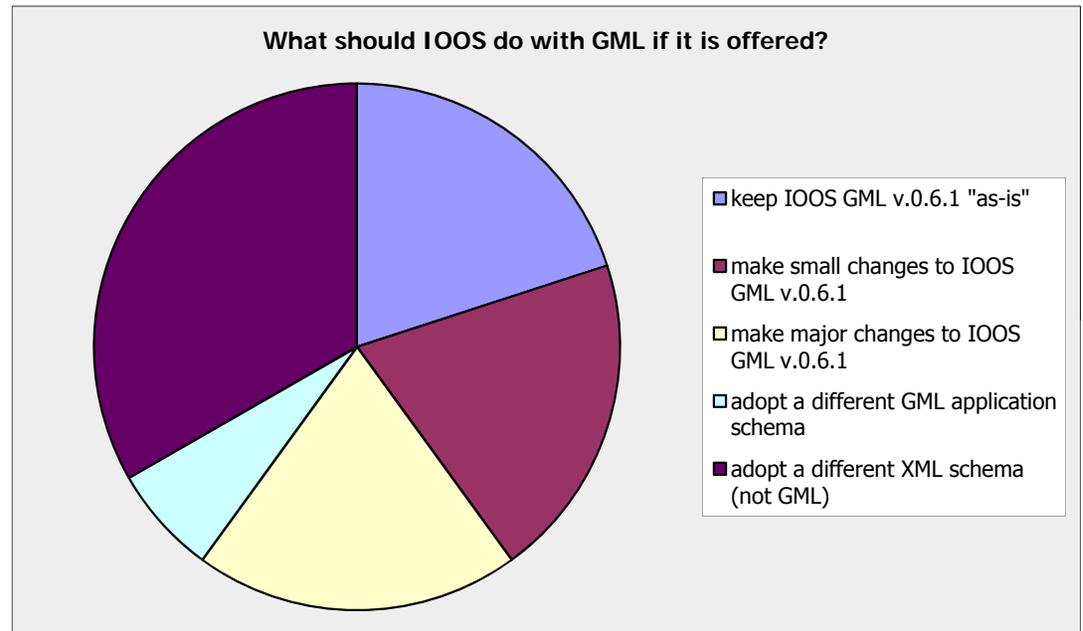
Answer Options	Response Percent	Response Count
Yes, GML only	5.3%	1
Yes, along with other formats	78.9%	15
No, GML should not be offered	15.8%	3
<i>answered question</i>		19
<i>skipped question</i>		4



DIF Encoding Evaluation

What should IOOS do with GML if it is offered?

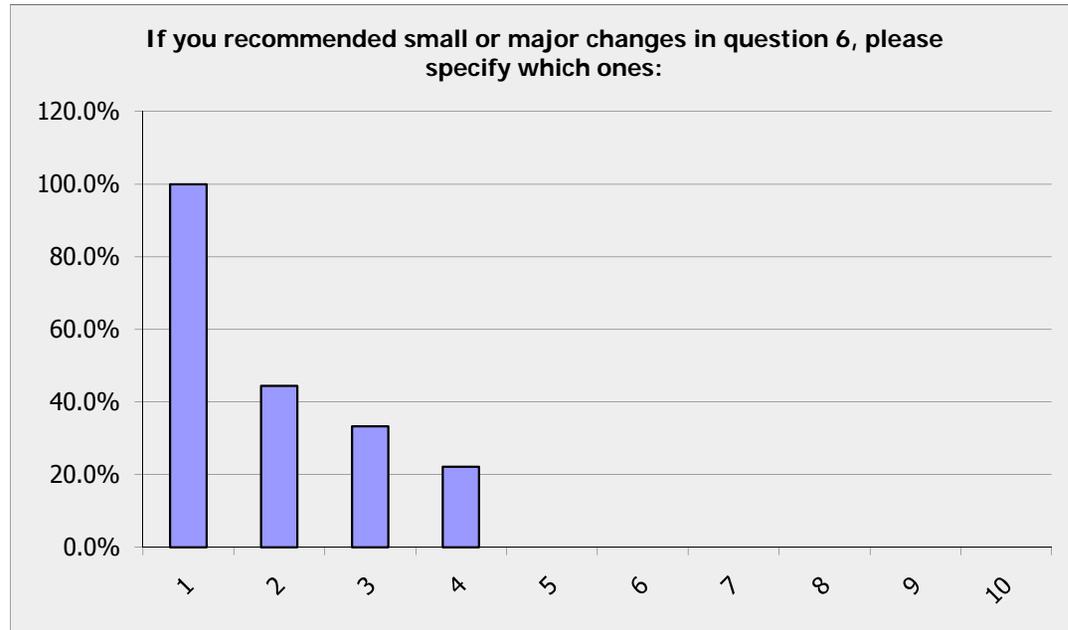
Answer Options	Response Percent	Response Count
keep IOOS GML v.0.6.1 "as-is"	20.0%	3
make small changes to IOOS GML v.0.6.1	20.0%	3
make major changes to IOOS GML v.0.6.1	20.0%	3
adopt a different GML application schema	6.7%	1
adopt a different XML schema (not GML)	33.3%	5
<i>answered question</i>		15
<i>skipped question</i>		8



DIF Encoding Evaluation

If you recommended small or major changes in question 6, please specify which ones:

Answer Options	Response Percent	Response Count
1	100.0%	9
2	44.4%	4
3	33.3%	3
4	22.2%	2
5	0.0%	0
6	0.0%	0
7	0.0%	0
8	0.0%	0
9	0.0%	0
10	0.0%	0
<i>answered question</i>		9
<i>skipped question</i>		14



DIF Encoding Evaluation

If you recommended small or major changes in question 6, please specify which ones:

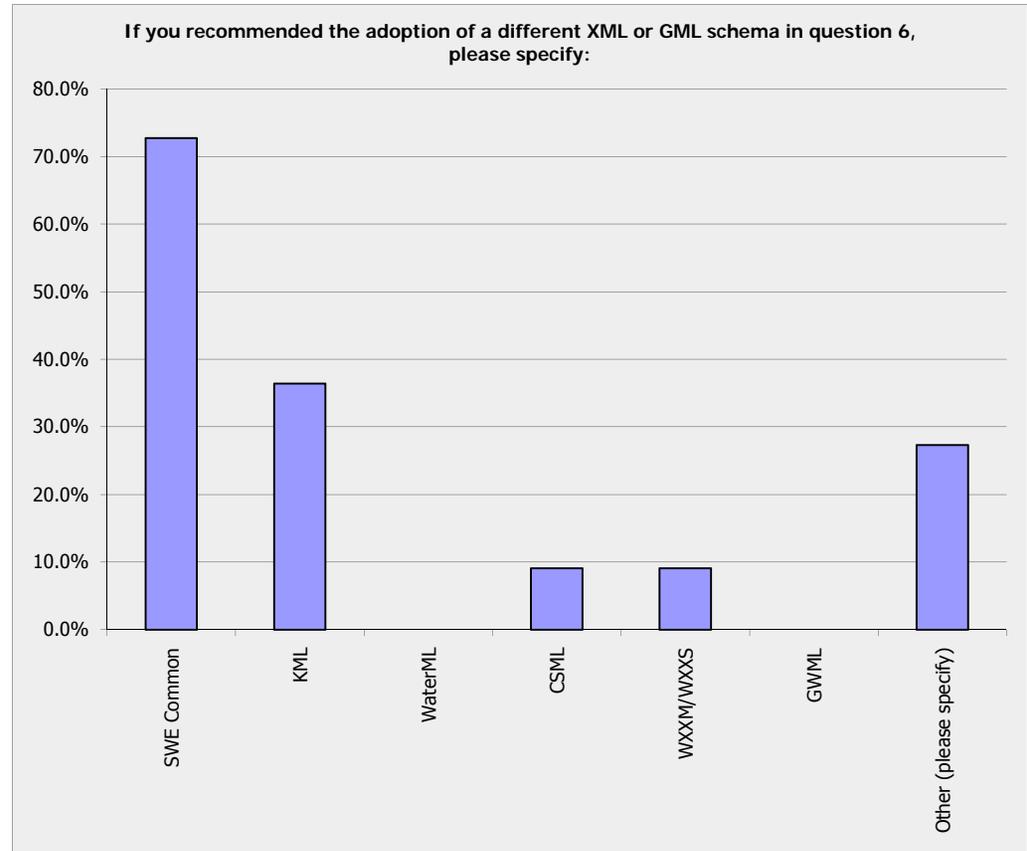
#	Response Date	
1	Dec 16, 2009 8:51 PM	1 create a collection of smaller less complex specialized schema
2	Dec 18, 2009 4:14 PM	1 if IOOS GML schema is mostly the same except for changes in the observation type(say water temperature versus salinity), don't articulate separate schemas per observation type, instead try to abstract the observation type as scalar or vector and minimize the number of schemas listed hopefully to just a handful that can be recycled and better documented
3	Dec 24, 2009 8:28 PM	1 The new SOS standard supposedly will specify SWE common. If this is about standards.....
4	Jan 7, 2010 5:35 PM	1 A smaller more compact subset of GML should be used
		2 Simplify the XML Schema such that not all OGC schemas get included.
		3 Data definition schema should only include what you need
5	Jan 7, 2010 6:06 PM	1 The URN issue needs to be clarified! Data submitted to NDBC by a non-federal station are not reflected in the URN and perhaps, it is about time to give up the urn:x-noaa and start the process of registering the ioos namespace.
6	Jan 7, 2010 6:16 PM	1 units vocabulary based on Unidata's udunits library which will also conform to CF standards
		2 transition to CF vocabulary for GetObservation requests, some existing names can be kept as aggregate groups of observations (Winds)
		3 In general, continue forward with development and implementation of the GML spec.
		4 Adopt other formats as reasonable, from Q#5
7	Jan 13, 2010 1:59 PM	1 I believe that the OGC's O&M schema should have been used for GetObservation responses. It is part of SWE and would be more compatible with international efforts including CSML which is aligning with O&M
8	Jan 13, 2010 6:04 PM	1 Specify uniform time format: NDBC - YYYY:MM:DDT:HH:MMZ; CO-OPS - YYYY:MM:DDT:HH:MM:SSZ
		2 GetObservation Request parameters are not uniform:CO-OPS requires 'version' parameter; NDBC does not require version parameter
		3 A clear hierarchy is lacking for the procedure and property elements. The current schema makes it difficult to construct GetObservations requests and it is difficult to determine the relationship between procedures and properties.
		4) NDBC parameter allowed values character capitalization does not match GetObservation, 'observedProperty' name
9	Jan 14, 2010 4:31 PM	1 Attempt to reduce verbosity if possible in the GML.
		2 If not, consider another option.

DIF Encoding Evaluation

If you recommended the adoption of a different XML or GML schema in question 6, please specify:

Answer Options	Response Percent	Response Count
SWE Common	72.7%	8
KML	36.4%	4
WaterML	0.0%	0
CSML	9.1%	1
WXXM/WXXS	9.1%	1
GWML	0.0%	0
Other (please specify)	27.3%	3
answered question		11
skipped question		12

Number	Response Date	Other (please specify)
1	Dec 17, 2009 4:15 PM	offer XSLTs to translate content among a base schema and these listed above
2	Jan 7, 2010 5:35 PM	A simplified version of the current schema.
3	Jan 15, 2010 6:34 PM	Adopt SWE Common encoding, which includes an XML encoding. Work with SWE Common WG to advance a standard encoding that can fit the needs for IOOS encoding.

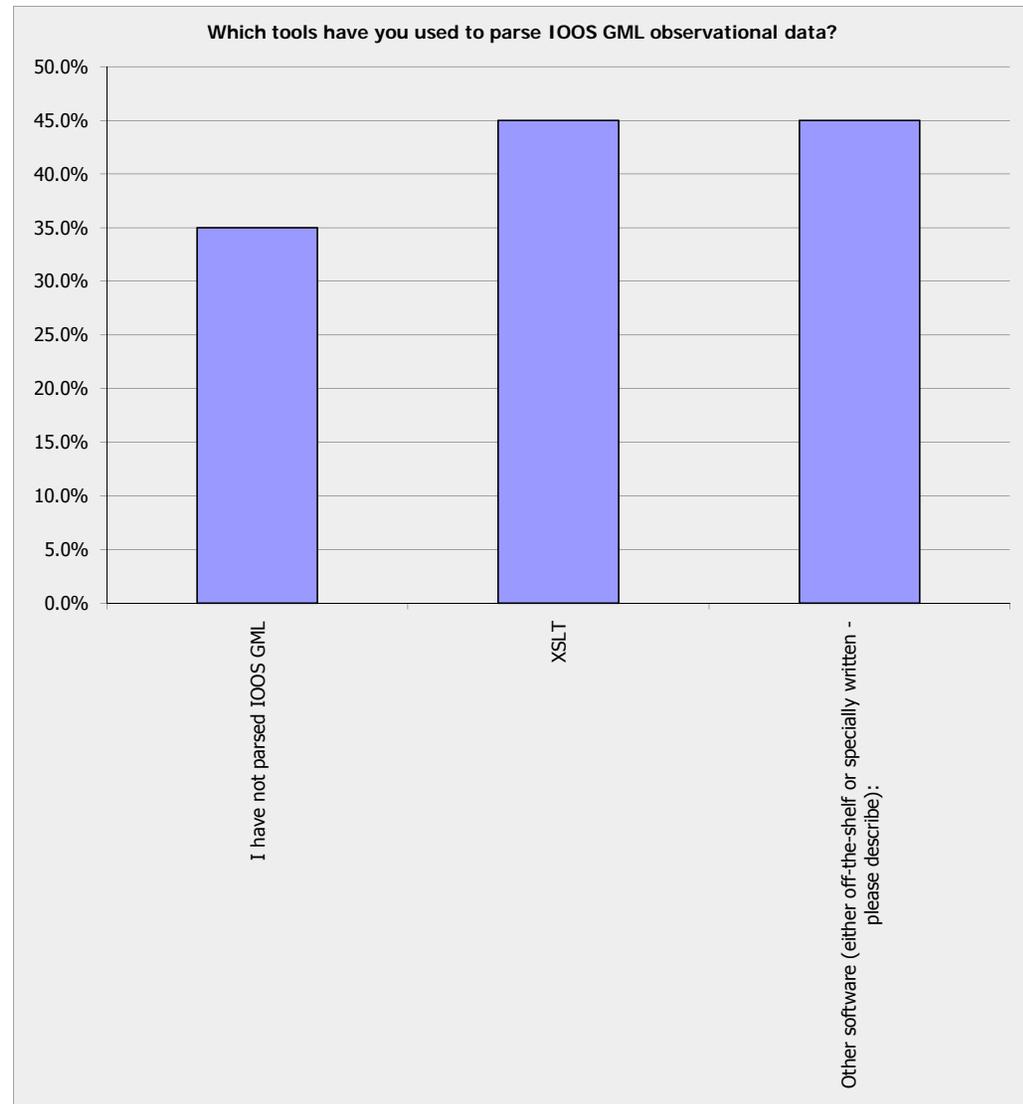


DIF Encoding Evaluation

Which tools have you used to parse IOOS GML observational data?

Answer Options	Response Percent	Response Count
I have not parsed IOOS GML	35.0%	7
XSLT	45.0%	9
Other software (either off-the-shelf or specially written - please describe):	45.0%	9
<i>answered question</i>		20
<i>skipped question</i>		3

Number	Response Date	Other software (either off-the-shelf or specially written - please describe):
1	Dec 16, 2009 8:51 PM	XMLSpy, Excel
2	Dec 16, 2009 10:04 PM	ERDDAP
3	Dec 18, 2009 4:14 PM	perl package XML::LibXML, uses XPath references
4	Dec 24, 2009 8:28 PM	We have our own parser
5	Jan 7, 2010 5:35 PM	Jaxb, Apache XML-Beans
6	Jan 7, 2010 6:16 PM	At present, we use the lxml python module to parse what is needed from the SOS services.
7	Jan 13, 2010 1:59 PM	LibXML C library API for which numerous bindings exist: Perl, Python, Ruby, Java, etc. The API includes support for XPath searches of XML documents.
8	Jan 13, 2010 6:04 PM	Wrote the SOS client for the Environmental Data Connector using Java JDOM and XPATH.
9	Jan 14, 2010 4:31 PM	JavaScript/DOM



DIF Encoding Evaluation

If you have any additional comments about IOOS SOS or GML implementation, please provide them here:

Answer Options	Response Count
	13
<i>answered question</i>	13
<i>skipped question</i>	10

Number	Response Date	Response Text
1	Dec 16, 2009 8:51 PM	Like other OGC specs, the SOS schema is loosely designed to accommodate a broad base, thus resulting in a difficult moving target for data consumers/vendors trying to build clients - thus reducing uptake. For a highly engineered (stable) back-end (sensor to shore) environment, the existing schema could be fine. I don't think either the SOS service or GML work well near the data delivery point to users.
2	Dec 16, 2009 10:04 PM	IOOS SOS is excessively complex. Instead of being a general solution to encoding a broad range of data types, IOOS SOS looks like a brittle (hence ever-changing) and complex solution to a few data types. It is unfortunate that IOOS SOS is different from the Oostethys group's work, which was intended to unify and standardize the use of SOS.

I do not see the need for providing one or more additional output formats. To do so may imply that there is something wrong with the one we've got. If there is, perhaps we should make it better, or adopt an existing one that is better. I imagine this comes up because there are a lot of people out there with clients that support ios gml and the translation is easy to do. I don't think those are good criteria for deciding on a schema for sos.

Perhaps it boils down to should the translation be done on the server or the client? I think that if we focus on getting one good standardized data dissemination tool then there are millions of programmers out there who will create the applications that can handle the ios gml conversion, the kml conversion as well as a myriad of other conversions that we don't even know about yet.

3 Dec 17, 2009 1:35 AM

We need to ask ourselves: "what is wrong with the one xml schema we already have?".

Possible answers:

- 1) there are few if any clients out there that support the current sos schema. But if it's a good schema certainly people will jump on the bandwagon and provide such clients.
- 2) the current sos schema is difficult to work with, clunky, overly verbose, not verbose enough. I.e. somehow deficient as it stands. Then we should fix it.
- 3) the current sos schema is limited in ways that other schemas are not. Then we should adopt a better schema.
- 4) the current schema will not provide a suitable platform for disseminating data in the future as say more complex data types are supported. Then we should adopt a better schema.

Offer the ascii tabulated data from a given station/time period in the formats that NDBC currently serves from its station pages:

4 Dec 17, 2009 4:15 PM

```
12 17 8:50 am E 21.4 25.3 8.2 7 6.2 - 29.97 +0.01 75.2 77.0 - - - -  
12 17 7:50 am ESE 21.4 25.3 7.9 8 6.3 - 29.97 +0.00 75.4 76.8 - - - -  
12 17 6:50 am E 21.4 25.3 7.5 8 6.3 - 29.96 -0.02 74.5 76.5 - - - -
```

rather than separate parameters delivered in separate requests.

For small amounts of recent data(say the past few days from an hourly reporting platform), I think the GeoJSON is more attractive than XML as a web/browser centric format due to more common javascript oriented processing and libraries.

5 Dec 18, 2009 4:14 PM

For larger amounts of data, XML or hierarchy structure 'gets in the way' and bloats the transmission or processing of data - in these cases I think its ok to collapse the data to CSV blocks or other compressed/binary formats, perhaps with XML metadata describing how the CSV or binary data attributes and how they should be processed.

(off the cuff thoughts as I am stuck on a long telcon ...)□

□

In our current generation of software we find ourselves in a transition between the older, well-established, file-based data management and a vision of service-oriented interchange in which "files" are increasingly unnecessary. A lesson to share from the OPeNDAP experiences of the last decade and a half: the transition process is a long struggle still has a long ways to go. Users still want "files" and their working environments still require them for many purposes (as well as working habits) today.□

□

In the discussions of XML here there is no mention of persistent formats ... of "files". That may leave many user needs of the next (say) 5-10 years unmet. Is there an implicit assumption that the DIF XML documents that are transmitted via SOS will also serve as persistent representations? Has thought been given to the significantly altered requirements that confront us when thinking of the XML as a persistent representation? Or is it assumed that every intense data user will be running their own private DBMS (with their own private DB schema) and ingest DIF info into that DB? Or alternatively is there a thought that (say) netCDF-CF should serve as a persistent format? (Has work been done to explore that?)□

6 Dec 18, 2009 8:51 PM □

-----□

□

The other element in the use of GML from SOS that is unclear to me -- again a "client" perspective -- is the degree to which there is (or is not) a data model the unifies the underlying the DIF encodings. For example, a piece of software that knows how to analyze time series may want to regard ADCP records as a series of time series at varying depths, or a model grid as time series at a 3D array of grid points. Is the expectation that the data modeling exists only *inside* of the application and that the application developer write individual blocks of code to parse the separate XML representations (or other) of data that it sees when reading time series, ADCPs, and grids? The Unidata and CSML communities have come together to address this issue -- much of which is only partially implemented or crystalized -- but there is momentum in that quarter. Not clear where the DIF XML may fit into that framework. Has in DIF anyone given thought to layering the Unidata CDM API concepts over the SOS requests and the GML encodings that DIF returns? □

□

Gotta get back to the telcon ... I hope this blurb makes sense.

Question 6 didn't allow for all possibilities of answers. For now, there are a group of services running, let's not disturb that, some minor tweaks. In the long run, if this is about standards, then the new SOS standard requires SWE common, so we should work towards that. As is, there are about 3 versions of SOS responses floating around, and it takes extra code to efficiently parse each. If it is about compatibility, then we should be looking at the underlying data model, and encode that data model different ways. The CSML and CDM data models are very close to each other, and if these could be represented in SWE or GML or netCDF or whatever, then at least we have common basic structures that are well understood across communities. At the last OGC meeting. at the MET Ocean DWG there were quite a few people mentioning that the entire data model in GML really isn't quite right for the problems faced by that community.□

7 Dec 24, 2009 8:28 PM □

□

I am still concerned about the speed of all of the SOS encodings for large datasets. SOS may work well for sensor to observatory, or for just the present data, but for large data requests by a user (say 20 years of hourly buoy data) the user can wait till the cow come home and the request still will not be complete. Are we saying there should be two different responses for small and large datasets? Is this good design? I don't think we can avoid this much longer.

- 8 **Jan 7, 2010 5:35 PM** The data definition schema should only include the elements needed to describe the data. The current schema includes schemas that includes schemas that include schemas until you get a parse tree that is huge, overly complex, and brittle. Most of the elements and definitions in the included schemas are not used. A data model/definition should only include what is needed and not a lot of useless fluff. XML documents based on schemas that are not overly verbose and complex are fast to parse and much easier to understand and maintain. This avoids what could be classed as Death by XML.
- 9 **Jan 7, 2010 6:06 PM** Installation instructions and most specially, updating implementations modules is lacking.
- 10 **Jan 7, 2010 6:16 PM** The same problem befalls any service be it SOS, WMS, WCS or DAP. In reality you can create a response that is too big. The WSDE group has kicked around two alternatives: compression and caching. Even if the SOS is replaced with a binary response like NetCDF, how is that different than the binary response now from DAP (XDR)? Another possible solution is server limits. For example: If a large request is made, say for 3000 records, and say the server has a limit of 500 records. The client should be able to systematically make 6 smaller requests based on limits that could be exposed in the GetCapabilities message. And maybe that is it, the regional and agency nodes will be small and fast and the OOI SOS service will be able to handle the large and bulky requests?
- 11 **Jan 13, 2010 5:46 PM** Achieving a balancing between offering flexibility and uniformity is a challenge and is a goal.
- 12 **Jan 13, 2010 6:04 PM** Server response construction is too slow and the resulting response to verbose to transmit. The schema is designed around returning the latest value, but providing for historical data access adds a difficult further requirement that is difficult to satisfy using the current schema.
- 13 **Jan 13, 2010 7:40 PM** Sorry I could not answer the remainder of the questions