

# TeamRoom Plus

IBM TeamRoom

This OPEN document will not be filed. It is being kept active.

## Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	01-Sep-10 (Wed)
Meeting Time	15:00 - 16:30

Created by Steve Hanson on 26-Aug-10

Last Modified by Steve Hanson on 01-Sep-10

## OGF DFDL Working Group Call , September 1 2010

### Agenda

**Prepare** for your meeting by describing the objectives (both immediate and long-term, if appropriate) of the meeting; and describe key details.

### Minutes

#### Meeting Minutes

**Reflect** on your meeting as you record all topics and issues discussed , and any tabled conversations . What went well, or what would you do differently next time? Document those so others can take advantage of your learning .

#### Open Grid Forum : Data Format Description Language Working Group

#### OGF DFDL Working Group Call , August 25-2010

##### Attendees

Stephanie Fetzner (IBM)

Steve Hanson (IBM)

Tim Kimber(IBM)

Bob McGrath (National Center for Supercomputing Applications)

##### Apologies

Mike Beckerle (Oco)

Alan Powell (IBM)

Suman Kalia (IBM)

Alejandro Rodriguez (National Center for Supercomputing Applications)

#### 1. Current Actions

[Updated Below](#)

#### 2. Twos complement integer lengths .

See new action 115 below.

#### 3. Validating a DFDL schema prior to DFDL processing

Not required by spec, so if it happens it must be careful when validating global objects .

This is not a spec issue, but is something for IBM implementation to watch out for.

NCSA implementation does not do this.  
Agreed that this was out of scope for WG call.

Meeting closed, 16:05

Next call Wednesday 8 September 2010 15:00 UK (10:00 ET)

### Create Action Items

Record the to-do's and individuals assigned by entering the appropriate information in the form below . Press the "Create Action Items" button to create specific to do's that can be tracked in the assignee's Work for Me views. " All Action Items will be tracked in the Action Items and Other Meeting Documents tab .

### Action Items and Other Meeting Documents

View: ResultDocs

Next action: 116

### Actions raised at this meeting

No	Action
115	Clarify allowed lengths for signed integer types when rep is binary integer (ie, two's complement) <a href="#">01/09: No technical reason to restrict lengths to 2^x bytes, could be odd, could be bits. But rare in practise so if we do relax, limit any core subset to 2^x bytes.</a>

### Current Actions :

No	Action
066	Investigate format for defining test cases 25/11:IBM to see if it is possible to publish its test case format . 04/12: no update ...

	<p>17/02: IBM is willing in principle to publish the test case format and some of the test cases. May need some time to build a 'compliance suite'</p> <p>24/03: No progress</p> <p>03/03: Discussions have been taking place on the subset of tests that will be provided.</p> <p>10/03: work is progressing</p> <p>17/03: work is progressing</p> <p>31/03: work is progressing</p> <p>14/04: And XML test case format has been defined and is being tested.</p> <p>21/04: Schema for TDML defined. Need to define how this and the test cases will be made public</p> <p>05/05: Work still progressing</p> <p>12/05: Work still progressing</p> <p>02/06: Work still progressing on technical and legal considerations</p> <p>...</p> <p>25/08: Will chase to allow Daffodil access to test cases. The WG should define how implementation confirm that they 'conform to DFDL v1'</p> <p>01/09: IBM still progressing the legal aspect. Intends to publish 100 or so tests as soon as it can, ahead of a full compliance suite.</p>
085	<p>ALL: publicise Public comments phase to ensure a good review..</p> <p>14/04: see minutes</p> <p>21/04: Press release, OMG and other standards bodies.</p> <p>05/05: Alan and Steve H have contacted other standards bodies. Will ask them to add comments on spec</p> <p>15/05: still no public comments</p> <p>02/06: No public comments</p> <p>16/06: Public comments period has ended with no external comments. Alan had posted changes made in draft 041. Steve suggested send a note to the WG highlighting these changes. Steve also suggested requesting an extension as other IBM groups may review. We discussed whether this was necessary as changes will need to be made during the implementation phase anyway. Alan to ask OGF what the process is for changes post public comment.</p> <p>23/06: Still no comments. Alan will contact OGF to understand the rest of the process.</p> <p>30/06: Alan has emailed Joel asking what the process is now public comment period is over and can we update the published version with WG updates. No response yet.</p> <p>07/07: No response. Alan will chase up</p> <p>14/07: No response from Joel. Sent email to Greg Newby by no response.</p> <p>21/07: Still no response.</p> <p>04/08: Joel has responded that it is up to the WG to decide if the changes are significant enough to need additional review. Alan to contact David Martin and Erwin Laure for guidance if we split the specification.</p> <p>11/08: Received a response from Joel that the WG can decide if a re- public review is necessary before becoming a 'proposed recommendation'. Alan responded that the WG agreed that a re-review was not necessary. The next stage is for OGF review committee to approve publication.</p> <p>11/08: Specification is now 'awaiting author changes' before being submitted to the OGF technical committee for approval as a 'proposed specification'.</p> <p>Alan would like to have the updated specification complete by Sept 10th. The WG needs to complete all actions by then or decide that they do not need to be included in this phase of the process.</p> <p>01/09: Alan and Steve have discussed and propose Sept 30th for completion of draft 43 and closure of all actions.</p>
099	<p>Splitting the specification in simpler sections.</p> <p>07/07: Steve sent a proposal but not discussed. Alan will arrange a separate call.</p> <p>14/07: Discussed Steve's proposal and Suman's and Alan's comments.</p> <p>Need to add choice, validation, facets.</p> <p>Also how does an implementation declare which subsets it supports. Suggested levels and/or profiles. Steve highlighted a problem when a DFDL schema from an implementation of just the core functions was moved to a full DFDL implementation what should happen about the missing properties. Does the full implementation need to be aware of subsets of functions? Should it raise a schema definition error for use of a function not in the subset.</p> <p>21/07: no progress</p>

	<p>04/08: Steve had updated proposed groups of function. (Subset_proposal_v2.ppt). We discussed whether its is better to have discrete sets of functions or expanding levels of function.</p> <p>Purpose of subsetting is:</p> <ol style="list-style-type: none"> <li>1. Allow simpler implementations. (main purpose)</li> <li>2. Simplify tooling</li> <li>3. Simplify specification.</li> </ol> <p>Steve to contact previous members of WG to check if we have the correct subsets</p> <p>11/08: Steve sent an email to previous members of the WG asking for opinions on splitting the specification. Bob McGrath from National Center For Supercomputing responded that they had implemented about 80% of the function. Alejandro will send a description of the function they have implemented.</p> <p>Action will be raised to track the Daffodil implementation</p> <p>11/08: not discussed</p> <p><a href="#">01/09: NCSA implementation description received. Making the unparser optional is a good idea (NCSA do not need one) . Work will progress on the subsets.</a></p>
101	<p><b>Semantics of 'fixed'</b></p> <p>21/07: Discussed whether not matching the 'fixed' value should be a validation error or processing error. Decided that for consistency it should be a validation error.</p> <p>It would be useful however to avoid having to duplication of facet information in an assert which could become unwieldy for, say, a large enumeration.</p> <p>Suggestions</p> <ul style="list-style-type: none"> <li>- a parser option that 'converted all validation errors to processing errors'</li> <li>- a dfdl expression function that 'applied all facets' or 'applied specific facet' to a particular element.</li> </ul> <p>Stephanie will produce some examples of how this could be used..</p> <p>04/08: Stephanie had produced examples but they were not discussed due to lack of time</p> <p>11/08: We started to discuss Stephanie's HIPPA example but ran out of time.</p> <p>25/08: Not discussed</p> <p><a href="#">01/09: Discuss next week</a></p>
107	<p><b>teston/testoff dfdl expression functions .</b></p> <p>Are these functions still needed. They were introduced to allow individual bits to be set in a byte. Steve to look at TLog and ISO 8583 formats that use existence flags to see if they are still required.</p> <p>04/08: Not discussed</p> <p>11/08: Not discussed</p> <p>25/08: Not discussed</p> <p><a href="#">01/09: Steve to progress by Sept 30th</a></p>
108	<p><b>dfdl:hidden</b></p> <p>There has been some discussion on whether the 'hidden' global group should be indicated in some way.</p> <p>04/08: A lively discussion. The specification is works as currently defined so whether changes need to be made to make tooling easier. There shouldn't be 'conventions' in particular tooling as they must be able to properly deal with schema from other tools that would not obey those conventions. Steve stated that it is often dangerous to hide too much from users when they can see they underlying schema. To be continued.</p> <p>25/08: there has been some offline discussions about simplifying how hidden elements are implemented. The proposal is</p> <ul style="list-style-type: none"> <li>• dfdl:hidden property on xs:element only</li> <li>• xs:minOccurs and xs:maxOccurs MUST be 0 when hidden</li> <li>• dfdl:minOccurs and dfdl:maxOccurs for hidden elements only.</li> <li>• An element is 'required' when dfdl:minOccurs &gt;0 and normal default processing occurs.</li> </ul> <p>The schema, without dfdl annotations, must match the info set so assumption is that non-DFDL tools, such as mappers, will ignore/not show elements with xs:minOccurs and xs:maxOccurs = '0'</p> <p><a href="#">01/09: The above proposal is flawed due to use of maxOccurs = 0 (this was identified back in 2008 hence current spec).</a></p> <p><a href="#">Bob confirmed that NCSA models use hidden in a big way, so punting hidden beyond 1.0 is not an option.</a></p> <p><a href="#">Two candidates:</a></p>

	<p>- As per spec but with syntactic improvements to make it clear that the two xs:sequences do not take any dfdl:sequence properties</p> <p>- Place a flag directly on a local element and force minOccurs to be 0. Simpler syntax but the semantic changes, as the element *could* be legally in the infoset, although a DFDL parser would never put it there.</p> <p>Steve will circulate the two proposals for next week.</p> <p>Bob to talk to Alejandro as the NCSA implementation is currently more flexible than the spec, allowing the groupref to point to a choice, and an elementref. Are these really needed?</p>
111	<p><b>Daffodil DFDL parser</b></p> <p>11/08: Bob and Alejandro described the new implementation that they have developed. It is a new code base and is not based on the Deffudle prototype. It is written in scala and implements approximately 80% of the features in the public comments draft of DFDL V1. Alejandro will send a list of the features not implemented.</p> <p>We discussed the scenarios that motivated the development which was to extract data from various sources and transform into canonical formats.</p> <p>Bob offered to make Daffodil available for the WG to assess the functionality. IBM WG members will get approval the company to allow them to receive Daffodil.</p> <p>Bob raised the question that if Daffodil becomes the public implementation of DFDL then we will need to work out how that would be funded and managed.</p> <p>It would be helpful if IBM test cases were available to Daffodil. IBM will investigate</p> <p>25/08: Alejandro had sent a list of the functions that he has implemented and Steve and responding indicating the extra functions he thought were essential.</p> <p>Since then Alejandro has implemented some of the missing functions, such as escape schemes, pre-defined variables, binary decimal numbers, etc, and will update his list.</p> <p>Bob is planning to make the parser available on the internet to allow testing.</p> <p>His organisation is being reorganised and he doesn't know what the priority of Daffodil will be so it is essential that we move quickly. It would help if IBM could indicate its support for Daffodil in some semi-formal way.</p> <p>01/09: Alejandro updating Daffodil to include escape schemes, unordered sequences and ignoreCase.</p> <p>Daffodil being placed under formal source control in anticipation of external release.</p> <p>Bob has a start October deadline to create a report on what has been done for his sponsors. It would be great if we could get Daffodil on the web and have run some IBM tests so it could be highlighted at OGF 30 at end October.</p>
112	<p><b>DFDL certification process</b></p> <p>25/08: Discussed how to certify DFDL implementations. Alan to investigate if OGF have a defined process.</p> <p>01/09: In progress, spec needs to state what conformance means, as part of this work</p>
113	<p><b>2. Regular Expressions .</b></p> <p>25/08: The DFDL regular expressions should provide lookahead and backreferences. Is the current regular expression language sufficient?</p> <p>Discussed two aspects:</p> <p>a. Is the XML regular expression language the correct one to use. Tim asked if DFDL needs to specify an language at all and should leave it to implementers to pick one. That would inhibit portability of schema.</p> <p>b. A regular expression property on an assert/discriminator as an alternative to the test expression. Either a DFDL expression or a regular expression could be specified but not both.</p> <p>01/09: There are many variations of regexp language, it seems wise to specify one that we know contains functions like lookahead, which makes it easy to say things like 'give me everything up to but not including x'. This rules out XML Schema and POSIX, it needs Perl 5 or Java.</p> <p>Tim to convince Steve (via example) that use of regexp in asserts is needed in 1.0.</p>
114	<p><b>3. OGF 30</b></p> <p>25/08: OGF30 takes place on October 25-29 in Brussels. Should we have a WG session?</p> <p>09/01: Given emergence of NCSA implementation and spec completion target of 30th Sept it makes sense to host a session at OGF 30.</p>

## Closed actions

No	Action
109	<p><b>dfdl:discriminator : the 'message' attribute</b></p> <p>From Tim:</p> <p>I remembered the reason why I thought this was a good idea.</p> <p>Consider the situation where someone is generating their DFDL schema from meta-data. The model is large, and consists of many references to global structures. Each global structure ( e.g. an HL7 segment ) is identified in a particular way. Sometimes the segment is required, sometimes it is not. Sometimes it occurs as a child of a choice group, and sometimes not. Regardless, it is highly likely that the segment will be identified in the same way wherever it occurs. A natural decision for the modeler would be to create a dfdl:discriminator on all references to the segment, even if the ref is not under a point of uncertainty. It's harmless, and it carries no performance penalty. If we disallow the "message" attribute, it will force the modeler to put in extra logic to work out whether the ref is under a POI, and generate an assert/discriminator as appropriate.</p> <p>I'd be interested to know what Steph thinks about this - I think I've heard her say that she sometimes uses discriminators where an assert would have done the job, just to maintain consistency throughout the model.</p> <p>04/08: not discussed.</p> <p>11/08: Not discussed</p> <p>25/08: Not discussed</p> <p>01/09: <b>Closed:</b> Conclusion was that this is genuinely useful, and has low implementation cost. Will add a 'message' attribute to dfdl:discriminator.</p>
110	<p><b>Semantics of newVariableInstance and setVariable</b></p> <p>11/08: what should a DFDL processor ( parser or serializer ) do when it cannot evaluate the expression in a newVariableInstance or setVariable annotation?</p> <p>Moving the setting of variable values after the element has been parsed just creates other problems. A new instance must be available to other expressions on the same component, and to the children of a group/element. So it cannot be left until the end of the element.</p> <p>On the other hand, there are clearly some types of setVariable / newVariableInstance annotations which *cannot* be evaluated until the element has been parsed.</p> <p>For the parser, it might be OK to</p> <ul style="list-style-type: none"> <li>- evaluate the expression when the component ( element or group ) is started</li> <li>- if it cannot be evaluated, add it to a list of annotations that must be processed at the end of the component</li> <li>- if in the mean time any other expressions attempt to access the variable that was being set/created then throw a processing error ( because the result will be undefined ). This will probably require the variable/instance to be placed into a 'not available' state until its expression is resolvable</li> </ul> <p>25/08: There was a brief discussion as IBM needs a resolution soon. Is it possible to restrict newVariableInstance to backward references only so remove the problem? setVariable must obviously be able to access the current value.</p> <p>01/09: <b>Closed:</b> Spec should distinguish newVariableInstance default value from setVariable value.</p> <p>For newVariableInstance default value, disallow downward references and references to self (must be usable from the point of declaration)</p> <p>For setVariable allow downward references and references to self, and always evaluate at end of component.</p> <p>(defineVariable default value should be same as newVariableInstance)</p>

## Work items:

No	Item	target version
005	Improvements on property descriptions	

0 1 2	Reordering the properties discussion: move representation earlier, improve flow of topics	
0 3 6	Update dfdl schema with change properties	ongoing
0 4 2	Mapping of the DFDL info set to XDM	none
0 7 0	Write DFDL primer	
0 7 1	Write test cases.	
0 8 3	Implement RFC2116	
1 0 9	Add 'message' attribute to dfdl:discriminator	43
1 1 0	Clarify expression limitations for defineVariable, newVariableInstance and setVariable	43
1 1 3	Be specific about regular expression syntax	43
1 0 8	Updates to hidden mechanism	43
9 9	Updates to reflect subsetting and unparser optionality	43
1 1 2	Define what conformance to spec means	43
1 1 5	Clarify allowed lengths for signed binary integers	43