# TeamRoom Plus

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

#### Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	19-Jul-11 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11 Last Modified by Steve Hanson on 20-Jul-11

## OGF DFDL Working Group Call, 19 July 2011

#### Agenda

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

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

#### OGF DFDL Working Group Call, 19 July 2011

#### 1. Spec issues from IBM (continued)

As implementation has progressed, a number of issues have been spotted in the DFDL 1.0 spec by the IBM implementation team. Remaining items for resolution.

13.6.Should textStandardDecimalSeparator ignored when the logical type is not decimal/float/double?

13.6.1.1. Erroneously mentions sigDigits in reference to the BNF for textNumberPattern.

13.6.1.1. Formatting. Uses terms 'minimum/maximum integer/fraction digits' but does not define them.

13.9. Should state that textBooleanTrue/FalseRep properties are used after trimming when parsing, and before padding when unparsing. If lengthKind is explicit or implicit and either textPadKind or textTrimKind is none then the properties must have the same length else schema definition error.

13.16. Nil literal character. When representation is binary, the encoding property is not used, which implies that a nil value must only contain DFDL hex literals (%rXX;)

13.16. Clarify nilKind 'literalCharacter' and 'literalValue' and whether nil test is applied before or after trimming.

16.1. Revise entire section.

16.2. State it is a processing error if the stop value is missing when parsing.

#### 2. Spec issues from Mike Beckerle

A number of issues have been spotted by Mike. Some have been resolved by action 139 below, but others remain. See separate e-mail.

#### Minutes

**IBM TeamRo** 

**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, 19 July 2011

#### Attendees

Mike Beckerle (Deloitte) Steve Hanson (IBM) Steph Fetzer (IBM) Ryan Farrell (US Navy Research Lab) Adam Fox (US Navy Research Lab) Suman Kalia (IBM)

#### Apologies

#### 1. US NRL tricky format

Bit-oriented format where repeating fixed length complex data contains a 'last item' indicator. All items need to go into the infoset including the last. If data was character oriented then lengthKind="pattern" would work using a regexp to grab the entire array. Similarly could use a discriminator that failed if not '1' or '0'. But as this is bit data, neither of these will work. Mike and Steve to think about how this can be modelled in DFDL. New **action 146** raised.

#### 2. Spec issues from IBM (continued)

13.6.Should textStandardDecimalSeparator ignored when the logical type is not decimal/float/double? Agreed that it should be ignored.

13.6.1.1. Erroneously mentions sigDigits in reference to the BNF for textNumberPattern. Agreed that this is in error and that sigDigits references should be removed.

13.6.1.1. Formatting. Uses terms 'minimum/maximum integer/fraction digits' but does not define them. Agreed that these phrases are undefined and it is better to state in terms of the pattern content.

13.9. Should state that textBooleanTrue/FalseRep properties are used after trimming when parsing, and before padding when unparsing. If lengthKind is explicit or implicit and either textPadKind or textTrimKind is none then the properties must have the same length else schema definition error. Agreed.

13.16. nilKind literalCharacter. When representation is binary, the encoding property is not used, which implies that a nil value must only contain DFDL hex literals (%rXX;) Disagreed. Encoding is still present, it applies to initiator and terminator. Also need to support COBOL MOVE SPACES example, where %SP; is the null character yet type is binary.

13.16. Clarify nilKind 'literalCharacter' and 'literalValue' and whether nil test is applied before or after trimming. New **action 147** raised to investigate. On the face of it, it would seem that for literalCharacter you would want to apply nil processing first, else the nil character can get trimmed away. But for literalValue, trimming first is desirable to allow a single nil value to apply to multiple different fixed length elements.

16.1. Revise entire section.

Agreed that this needs revising. Handle under action 140.

16.2. State it is a processing error if the stop value is missing when parsing. Agreed this is a processing error.

Errata document will be updated.

#### 2. Spec issues from Mike Beckerle

5.2.2. Last paragraph starting "It is a processing error when a fixed-length string is found to have a number

of characters not equal to the fixed number". Agreed that this is not correct. Will be removed as part of action 139 updates

9.2 DFDL Syntax Grammar change to introduce concept of EnclosedItem. Disagreed. Only make this change if it is found to be helpful in the context of action 140.

12.1.1. What does the note after Table 14 mean. "Specifying the implicit alignment in bits does not imply that dfdl:lengthUnits 'bits' can be specified for all simple types". Agreed that clarification is needed. It is really saying that alignmentUnits and lengthUnits are independent and have their own rules for when they are applicable.

12.3. One line descriptions of 'delimited' and 'endOfParent' are not worded correctly in the property description of lengthKind. Agreed that the wording for delimited and endOfParent can be improved.

Errata document will be updated.

Meeting closed 16:00 UK

Next call Tues 26th 15:00 UK

## 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: 148

Actions raised at this meeting

No	Action
146	Model NRL's bit-oriented format where complex repeating element carries a 'last' indicator
	(Steve, Mike)
	19/07: NRL to send xsd and example of format.
147	Clarify the rules when for padding /trimming and nil handling interact (Steve, Mike) 19/07: It would seem that for literalCharacter you would want to applly nil processing first, else the nil character can get trimmed away. But for literalValue, trimming first is desirable to allow a single nil value to apply to multiple different fixed length elements. There is an inconsistency here, but is that ok?

# Current Actions :

	looked at the later lessons, and noted that they are more direct compared to the more work earlier lessons.			
	28/06: On hold.			
	19/07: On hold			
24	DFDL web content on OGF standards pages (Steve, Bob)			
	13/10: no progress			
	10/11: no progress			
	17/11: Alan has looked at the OGF web pages and there aren't many standards listed. Some of the links point to very short primers rather than the specification			
	08/12: Alan to produce some information to be ready for when spec is approved. Still no word about is it was discussed/approved at OGF meeting			
	15/12: no progress 22/12: Steve has developed a summary web page for DFDL which will be sent to OGF whe spec is approved.			
	12/01: Not heard from Joel about updated OGF pages Alan to chase. Will also track other site updates: Wikipedia, IBM developerworks etc. 19/01: Still no response from Joel.			
	Other web site that need updating - IBM virtual XML			
	- Defuddle - Wikipedia			
	- Need google trawl for others Also need to make spec and tutorials more accessible on the web, eg in pdf and/or html			
	format. 26/01: Still not heard from Joel about OGF web pages.			
	PDF versions of the Specification and tutorials have been uploaded to gridforge.			
	02/02: A DFDL web page is available at www.ogf.org/dfdl. We need to update the IBM virtual XML and MCSA Defuddle pages. Will ask Mike Beckerle to update his DFDL page. 09/02: Wikipedia DFDL page is available.			
	23/02: All sites except Defuddle have been updated. 02/03: NCSA web pages have been updated. The DFDL WG home page needs updating			
	and should provide links to spec and tutorials. 09/03: Steve has updated the DFDL WG home page and is in contact with Edinburgh University to update an old DFDL presentation. Would like to have a separate DFDL tutoria			
	page to link to the individual lessons. 30/03: Bob will chase the update of the old Defuddle web pages. IBM investigating			
	conversion of spec and tutorials from pdf to html for usability from browsers. 13/04: Bob still sorting out the Defuddle updates. IBM work to convert the spec to html has started.			
	04/05: Jim Myers has updated the sourceforge download page but not the Defuddle home/overview pages, Bob will chase. IBM has converted the spec to html pages, needs			
	some tidying before being made public. 18/05: IBM aims to publish the web version by end June. Steve checking with OGF whether the web spec is a 'derivative work' in terms of the copyright notice, or can be considered and			
	actual copy of the spec. 01/06: Still waiting on status of web spec from OGF			
	08/06: OGF have come back to us and said that it is ok to create a web version of the spec If it is unchanged then it is a copy, otherwise it is a derivative work. Either way, the copyrig covers this so there is not a problem. The web version is looking good, some minor tidy-up			
	changes needed where formatting is not quite right. When IBM is happy with it, Steve will circulate to the WG for review.			
	15/06: Ongoing 28/06: Steve needs to take a final look at all the pages, get any problems fixed, then distribute to the WG			
	05/07: Reviewed, some editorial changes needed, and a problem with the table numbers getting reset across pages. When these are fixed Steve will distribute. 12/07: Editorial changes made, fixing the table numbers and also XML indentation in			

132	19/07: Work has started to address the two issues noted on 12/07. Publishing DFDL xsd (Suman)
	08/12: Agreed that it should be made available. Suman has started the approval process in
	IBM
	15/12: no progress
	22/12: no update
	12/01: Suman is getting approval from IBM to publish.
	19/01: Waiting to get IBM approval to publish
	26/01: no update 02/02: No update
	09/02: no update
	23/02: no update
	02/03: Suman is working through the IBM process to permit publication. There was discussion about what licence the XSD would be published under and how that would effecuse in products. Suman to investigate
	09/03: No update 30/03: Suman has sent information to IBM legal. Reminded him about the license issue.
	13/04: No update.
	04/05: IBM has permission to release the DFDL model xsds to WG members only, Suman has a couple more changes to make and will send to Steve for review. License clarification needed.
	18/05: Awaiting response from IBM legal. Suman will send Steve the model xsd for review 01/06:
	08/06. Awaiting response from IBM legal. Suman will send Steve the model xsd for review 15/06: Steve has received the xsds (there are three of them) and will review.
	28/06: Not reviewed yet
	05/07: Need to fully understand what WG members are able to do with it. The real
	usefulness is in other implementers being able to use the xsds to validate DFDL xsds so the
	license needs to reflect this.
	12/07: No further progress 19/07: Suman will talk to IBM legal and make it clear that the license must allow users to
	actively use the xsds.
133	Make a set of default formats available (Suman)
	19/01: Suman expects some default formats to be ready by Feb 9th. Will need approval to publish
	26/01: Stephanie sent the defaults used by test cases to Suman 02/02: no update
	09/02: no update
	23/02: no update
	02/03: Same as 132
	09/03: No update. Same license issue apply though. 13/04: No update.
	04/05: IBM will make one default format available. Suman is working through the IBM process to permit publication.
	18/05: Awaiting response from IBM legal.
	01/06:
	08/06: Awaiting response from IBM legal. IBM also want to prove that the default format hat the properties sensibly defined to plans to include in internal testing. 15/06: No change.
	 19/07: 19/07: Suman will talk to IBM legal and make it clear that the license must allow users to actively use the format xsd(s).
136	Arrays with missing elements (Steve)
	There is a problem when there are empty/missing array elements with an index greater that minOccurs. For example:
	xs:element name"array" minOccurs=0, maxOccurs=10 lengthKind='delimited'
	Datastream: ,,,value3,value4

	Infoset will contain:			
	array[1] = value3 array[2] = value4			
	This is because elements with an index greater than minOccurs are optional and so do not get defaulted.			
	Unparsing this infoset will produce:			
	Datastream: value3,value4			
	You could make the empty space (%ES;) the nil value which will work for simple elements but not for complex.			
	Infoset will contain: array[1] = nil array[2] = nil array[3] = value3 array[4] = value4			
	23/02: Discussed options 1. Changed definition of required for arrays to be 'required up to the last instance in the data stream of the array'			
	<ul> <li>2. add index to the element info item</li> <li>Steve to investigate if XDM uses an index.</li> <li>02/03: No progress</li> </ul>			
	09/03: No update 30/03: Stephanie recognised the issue from IBM's WTX. Here, the solution was to provide an option so that the user explicitly chose whether the position in the array was significant, as it is not always and on output some users do not want defaults or blanks to appear. Not yet resolved.			
	13/04: Steve has verified that XDM does not carry index information, but will check with IBM's W3C rep to see if that has ever been expressed as a requirement. Steve also said that the area of defaulting missing required elements on parsing (especially complex elements) is one that the IBM implementation team has raised some concerns about, so Steve and the team are looking at this area again. It is possible that the spec will change to clarify behaviour, and so this action should be used to cover this work. 04/05: In progress.			
	18/05: In progress, at minimum some clarification to the spec is needed 01/06: Discussed to bring Mike up to speed. Solution will depend on other spec discussions 08/06: Still under discussion			
	15/06: Will come back to this after action 140 resolved 28/06: As above			
	 19/07: As above			
140	Spec issue: Parsing: 'missing' v 'empty', role of initiators , default values (All) 01/06: See minutes.			
	08/06: Still under discussion. Tim has sent Mike a selection of data formats to guide the discussion. 15/06: Not discussed - an extra call has been scheduled to go through this.			
	28/06: A series of extra calls are being held between Mike, Steve, Tim and Steph. 05/07: Next extra call is Wed 6th July - Steve to send invite 12/07: Two more calls held. Next call is Wed 13th July.			
141	19/07: More calls held, next call is Fri 22nd July.         Should text number exponent , infinity and Nan rep properties be lists ? (IBM)			
	28/6: There is certainly a requirement for DFDL to handle multiple reps for these properties If ignoreCase is not an option due to ICU then should these allow a list? 05/07: In progress with IBM.			
	12/07: Still with IBM 19/07: Still with IBM.			

145	Provide a 'dispatch' way of discriminating a choice for better performance of the
	envelope/payload use case (Steve, Mike)
	12/7: See minutes. Need to choose a proposal and flesh out.
	19/07: Waiting for proposals

## **Closed actions**

No	No Action			
-				
No 111	<ul> <li>Daffodil DFDL parser (Bob, Joe)</li> <li>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.</li> <li>We discussed the scenarios that motivated the development which was to extract data from various sources and transform into canonical formats.</li> <li>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.</li> <li>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.</li> <li>It would be helpful if IBM test cases were available to Daffodil. IBM will investigate 25/08: Alejandro had sent a list of the functions that he has implemented and Steve responded indicating the extra functions he thought were essential.</li> </ul>			
	schemes, pre-defined variables, binary decimal numbers, etc, and will update his list. Bob is planning to make the parser available on the internet to allow testing. His organisation is being reorganised and he doesn't know what the priority of Daffodill 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. 01/09: Alejandro updating Daffodil to include escape schemes, unordered sequences and			
	ignoreCase. Daffodil being placed under formal source control in anticipation of external release. 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. 08/09: Alejandro is marking up Spec draft 42 to indicate which features Daffodil implement. Bob expects Daffodil to be available on the web soon.			
	15/09: Alejandro had indicated in the specification which functions were implemented in Daffodill. Steve had reviewed and identified which function need to be implemented and which could be considered optional (see action 099). Alejandro is implementing the missing core functions. There was some discussion about the limitations on unordered groups. (stop value and expression not supported). It was agreed that it should be a schema definition error if dfdl:occursCountKind is 'stopValue' on any element within an unordered sequence and a floating element. 22/09: not discussed			
	<ul> <li>29/09: not discussed</li> <li>06/10: Alejandro has left NCSA. Bob is making the case for continuing and having a replacement. Bob to agree with Steve what can be said at OGF30.</li> <li>13/10: Bob still progressing project funding and making Daffodil publically available .</li> <li>10/11: NCSA internal &amp; sponsor (US National Archive in Washington DC - Electronic Records Administration) reviews passed. NCSA have new resource allocated - Joe Futrelle. Bob has started open source paperwork. ETA end December.</li> <li>17/11: Joe has started coming up to speed with Daffodil. Bob is waiting for signoff from the university to open source the code.</li> <li>24/11: not discussed</li> </ul>			
	<ul> <li>01/12: Joe is becoming familiar with the code. Waiting for university to give permission to make available.</li> <li>08/12: Disclosure process in progressing. Bob has permission to make the web page available so hopefully that will be soon.</li> <li>The US National Archive have an article on their website about DFDL. Unfortunately it is a little out of date.</li> </ul>			
1				

15/12: Hope to get permission to make the source code available within days. Web demo page will be available when resources have been identified.

22/12: An internal demonstration web page has been made available. Initial testing highlight some problems. Joe investigating.

12/01: Web site has been move to a public URL <u>http://daffodil.ncsa.uiuc.edu/</u>. Joe thinks there are multiple problems that will show up when he can run the test cases.

19/01: Joe has had little time to work on Daffodil. Without significant research into the code it is difficult to know how compliant daffodil is. Discussed if there are ways the WG could help. Suggestions

- Help review the code

- Review the Daffodil test cases for correctness

- Simplify the IBM test cases to be 'unit tests'

Alan to ask for permission to review test cases and possibly simplifying the IBM test cases 26/01: Joe has ot had any time to work on Daffodil. We briefly discuessed if it was possible to use XSLT to transform test cases so they didn't use defaulting but it didn't seem possible and effort would be better targetted at the Daffodil code.

Alan had asked for permission to view test cases and subsequently heard that it has been given.

02/02: Joe will send the Daffodil test cases for Alan to review.

09/02: Alan has received the test cases and will feed back comments. Joe is looking at support for the default format

16/02: Alan has corrected some of the daffodil test cases and identified common problems 23/02: No update

02/03: No progress

09/03: No progress

30/03: From Joe: "The current status of my work is that, having identified issues with the schema parser, I am now going to attempt to independently assess the processors. That is: I am going to hand-configure the processors in a series of tests so that I can see what they can parse and what they can't. In the process I may be able to refactor them so that they are less closely coupled to the schema parser, setting the stage for a new schema parser implementation."

13/04: Joe has analysed the schema parser part of Daffodil - it is looking for annotations in wrong place, scoping are rules broken, it is using wrong property names. Needs a rewrite. But the parser itself is in better shape. Unfortunately Joe is leaving NCSA in 4-5 weeks time. Net is that any Daffodil on the web will be a partial, non-conformant implementation. Must make it very clear that this is the case. Also unlikely that Joe will be replaced. Joe will see whether there is any chance to work on DFDL in his new job.

04/05: Bob will create some white papers on what to do when implementing a DFDL parser, on the state of the current Daffodil implementation, and on lessons learned. The former is of particular interest as the IBM implementation is throwing up some interesting design points and a couple of difficulties in the area of empty/missing/nil/default processing.

18/05: No update

01/06: No update

08/06: No update

15/06: No change in status. Bob wants to spread the new wider to find contributors to Daffodil, as NCSA does not have the qualified staff available, and no new hires are likely. US NSA have shown an interest in contributing. Steve has invited them to future DFDL WG calls.

28/06: No update

05/07: Adam Fox (US Navy) has been looking at the Daffodil parser, to see where the holes are in respect to the formats they are interested in. Alonza Mumford (US NCSC) will be doing the same this week for his formats. Need holistic view from the various US DoD projects as to the missing function. DoD have also suggested setting up a community source project so that any future changes to Daffodil parser are visible to the wider community.

12/07: Update from Mike. NCSA and NSA discussing how to complete the Daffodil parser in a timely manner.

19/07: Closed. US NRL now actively looking at Daffodil.

**144 Create errata document to track spec issues (Steve)** 05/07: Will be kept in grid forge and publicised on OGF and Wikipedia pages. This replaces the errata list at the foot of the minutes, which will be removed when the errata document is issued. 12/07: GWD Errata document created here: http://forge.gridforum.org/sf/go/doc16280?nav=1 . Work list updated below. 19/07: Closed. Steve has updated the DFDL Wikipedia.page.

## **Deferred actions**

No	Action
129	Press release to publicise DFDL (Steve) Steve is pulling together a press release at IBM. Want to include as many contributors and interested parties as possible.NCSA are keen to be included. Also likely that US National Archive will want to be included. Mike has indicated OCO are too. 17/11: no progress
	 08/12: Still no response from IBM press office 15/12: no progress
	 09/03: No progress 30/03: Making this action deferred until IBM is in a position to say something more concrete about any implementation.
131	Transformation of DFDL properties to a canonical form (Joe) 08/12: Joe has produced a XSLT to transform a DFDL schema to a canonical element form. When tested it should be made available on the WG gidforge site. 15/12: Alan tested against test dfdl schema which worked correctly (after fixing some errors in the schema) 22/12: no update
	<ul> <li>12/01: Joe has some defects to fix before making available on gridforge.</li> <li>19/01: There is a difficult problem to solve before Joe make the style sheet public</li> <li>26/01: Working on problems</li> <li>02/02: no progress</li> <li>09/02: As it wasn't a simple as exoected this will be treated as a low priority action</li> </ul>
	23/02: Low prioity 09/03: Low priority 30/03: Deferring for now
066	Investigate format for defining test cases (AII) 25/11:IBM to see if it is possible to publish its test case format. 04/12: no update
	 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' 24/03: No progress
	03/03: Discussions have been taking place on the subset of tests that will be provided. 10/03: work is progressing 17/03: work is progressing
	<ul> <li>31/03: work is progressing</li> <li>14/04: And XML test case format has been defined and is being tested.</li> <li>21/04. Schema for TDML defined. Need to define how this and the test cases will be made public</li> </ul>
	05/05: Work still progressing 12/05: Work still progressing 02/06: Work still progressing on technical and legal considerations
	 25/08: Will chase to allow Daffodil access to test cases. The WG should define how implementation confirm that they 'conform to DFDL v1' 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. 08/09: IBM still progressing
	<ul> <li>15/09: IBM still progressing, expect tests to be available within a few weeks</li> <li>22/09: IBM still progressing, expect tests to be available within a few weeks</li> <li>29/09:Test cases are being prepared.</li> <li>06/10: Some test cases should be available next week. Steve would like to be able to show</li> </ul>

## Work items:

	No	Item	Owner	Target	Status
Γ	043	Track errata list for 1.0 of the spec. http://forge.gridforum.org/sf/go/doc16280?nav=1	Steve	N/A	Ongoing
	044	Incorporate errata list into DFDL spec.	Steve	TBD	

© Copyright IBM Corp. 1998, 2007 All Rights Reserved