# TeamRoom Plus

IBM TeamRo

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

**Meeting about Meetings\OGF**

| | |
|---|---|
| Project | DFDL 1.0 |
| Meeting Date | 30-Aug-11 (Tues) |
| Meeting Time | 15:00 - 16:00 |

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 26-Aug-11

**OGF DFDL Working Group Call , 30 August 2011**

Agenda

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

**1. Further spec issues from IBM**
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.

13.6.1.2. This section is called "dfdl:textNumberPattern for dfdl:textNumberRep 'zoned' " but it is not entirely stand-alone, as it relies on a lot of common information from 13.6.1.1 "dfdl:textNumberPattern for dfdl:textNumberRep 'standard' ". But the words don't actually say this, which is leading to questions from IBM implementors. Proposed that 13.6.1.2 should formally list the sub-sections of 13.6.1.1 that apply, etc.

23.3. Expression syntax. Clarify that DFDL expression syntax "{}" is invalid, as it results in an empty XPath 2.0 expression, which is not legal. So for a property like dfdl:separator, "{}" is not the same as "" (empty string).

11. Spec does not definitively list which binary reps are subject to byteOrder. Assumed to be all Numbers and Calendars with representation binary, excluding floats. So that's binary integers, packed decimals, BCD, binary seconds and binary milliseconds.

13.11.1. When parsing an xs:date or xs:datetime, if a `calendarPattern` doesn't specify some parts (say, `calendarPattern`="MM") what values should be used? IBM implementation has assumed Unix epoch which is 1970-01-01T00:00:00.000GMT. An alternative is 0000-01-01T00:00:00.000GMT.

13.6. textStandardZeroRep. An earlier errata had limited the DFDL literals that several of the text number properties can take, including textStandardZeroRep. This included prohibiting WSP, WSP+, WSP*. Does this limit the usability of a known scenario where spaces is used when a value is zero?

**2. Validation of DFDL schemas**
Section 2.2 says "All conforming DFDL processors must detect all schema definition errors..".
That's a big task for an implementation.
- What if the implementation only implements a parser, should it validate unparsing only properties?
- What if the implementation does not implement optional features?
- Is this true even if the schema definition error is not encountered when parsing/unparsing the actual data?
- What is the recommendation for global elements/groups, where properties may be completed by element/group refs?
- What is the recommendation for simple types, where properties are combined with the using element?
- What if a benign but non-applicable property is explicitly supplied on an object?

3. AOB

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 .

**Attendees**
Steve Hanson (IBM)

**Apologies**

**Meeting closed**

**Next call**
Tues 6th Sep 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: **152**

## Actions raised at this meeting

| No | Action |
|---|---|
| | |
| | |

## Current Actions :

| No | Action |
|---|---|
| 123 | **DFDL tutorial (Steve)**<br>13/10: Draft of first 3 chapters has been written and will be distributed to WG<br>10/11: Posted to grid forge here (http://forge.gridforum.org/sf/go/doc16106?nav=1), work continuing at IBM to define a standard example-based chapter framework and to author additional chapters. Contributors welcome!<br>17/11: Steve, Stephanie and Alan had a meeting to discuss the best structure for the tutorial and decide which examples to use throughout. The meeting raised more questions. Further discussions will be held.<br>24:11: The list of topics to be covered in the remaining lessons has been produced and a lesson template. Alan will write lesson 4<br>01/12: Alan has started lesson 4 which covers fixed and variable fields and arrays.<br>08/12: Alan has almost completed lesson 4. Will send out for review.<br>15/12: First draft of lesson 4 is available for review. Alan to send to Bob and Joe.<br>22/12: Alan has distributed drafts for tutorials on Basic Structure and Optional/Repeating elements. Please review<br>12/01: Alan distributed a tutorial for choices and updated the others. Alan and Steve reviewed them and updated versions will be sent soon. Should start on the 'representation' tutorials soon.<br>19/01: The tutorials for basic structure, optional/arrays and choices have be updated. Please review. The tutorial for text elements should be available soon.<br>26/01: No comments received about 3 tuorials distributed last week. Alan is still working on Text representation.<br>02/02: Steve has sent comments on three tutorials. Alan to send updated versions by the end of the week. Alan has also distributed the first part of the tutorial on text representation and would like feedback.<br>09/02: Steve had reviewed tutorials 3,4,5 and updated versions have been distributed. Joe reviewed lesson on text elements.<br>Main points. Using 'represented as text' is confusing. Examples are too cluttered. Suggest simple targeted examples but still build up to final complete schema<br>23/02: New versions distributed and Steve has commented.<br>02/03: Alan has published the final versions of tutorials 4,5,6 and is working on text respresentations. There was some discussion about the detail that needs to be covered. Should limit it to 'common usage' and refer to the spec for details of edge cases.<br>09/03: Alan distributed an update to the text tutorial. Please review.<br>30/03: Steve has spent half a day tidying up lessons 1 to 6 and has uploaded them as pdfs to gridforge. They are now more coherent, and many inconsistencies and errors fixed. Ownership of draft lessons (text properties, binary properties, advanced features) has been passed to Steve. Also need to make a schema available for the examples.<br>13/04: Steve is working on the text properties tutorial.<br>04/05: No progress<br>18/05: No progress<br>01/06: No progress<br>08/06: No progress<br>15/06: This is on hold until Steve clears up spec issues and other workload. Steph has looked at the later lessons, and noted that they are more direct compared to the more wordy earlier lessons.<br>28/06: On hold.<br>...<br>26/07: On hold<br>16/08: On hold |

| | |
|---|---|
| | 23/08: On hold |
| 124 | **DFDL web content on OGF standards pages (Steve, Bob)**<br>13/10: no progress<br>10/11: no progress<br>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<br>...<br>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<br>15/12: no progress<br>22/12: Steve has developed a summary web page for DFDL which will be sent to OGF when spec is approved.<br>12/01: Not heard from Joel about updated OGF pages Alan to chase.<br>Will also track other site updates: Wikipedia, IBM developerworks etc.<br>19/01: Still no response from Joel.<br>Other web site that need updating<br>- IBM virtual XML<br>- Defuddle<br>- Wikipedia<br>- Need google trawl for others<br>Also need to make spec and tutorials more accessible on the web, eg in pdf and/or html format.<br>26/01: Still not heard from Joel about OGF web pages.<br>PDF versions of the Specification and tutorials have been uploaded to gridforge.<br>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.<br>09/02: Wikipedia DFDL page is available.<br>23/02: All sites except Defuddle have been updated.<br>02/03: NCSA web pages have been updated. The DFDL WG home page needs updating and should provide links to spec and tutorials.<br>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 tutorial page to link to the individual lessons.<br>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.<br>13/04: Bob still sorting out the Defuddle updates. IBM work to convert the spec to html has started.<br>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.<br>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 an actual copy of the spec.<br>01/06: Still waiting on status of web spec from OGF<br>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 copyright 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.<br>15/06: Ongoing<br>28/06: Steve needs to take a final look at all the pages, get any problems fixed, then distribute to the WG<br>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.<br>12/07: Editorial changes made, fixing the table numbers and also XML indentation in examples.<br>19/07: Work has started to address the two issues noted on 12/07.<br>26/07. Changes made, Steve will distribute to the WG prior to posting on the OGF web site and updating the Wikipedia link.<br>16/08: Steve has incorporated comments from WG and sent off to OGF standards council. Positive feedback from Alan Sill, awaiting final response. |

| | |
|---|---|
| | 23/08: OGF are using DFDL's web spec as an exemplar and have asked how it was created. Steve has forwarded details. OGF will add DFDL web spec to the DFDL pages on OGF we site. |
| 132 | **Publishing DFDL xsd (Suman)**<br>08/12: Agreed that it should be made available. Suman has started the approval process in IBM<br>15/12: no progress<br>22/12: no update<br>12/01: Suman is getting approval from IBM to publish.<br>19/01: Waiting to get IBM approval to publish<br>26/01: no update<br>02/02: No update<br>09/02: no update<br>23/02: no update<br>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 effect use in products. Suman to investigate<br>09/03: No update<br>30/03: Suman has sent information to IBM legal. Reminded him about the license issue.<br>13/04: No update.<br>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.<br>18/05: Awaiting response from IBM legal. Suman will send Steve the model xsd for review.<br>01/06:<br>08/06. Awaiting response from IBM legal. Suman will send Steve the model xsd for review.<br>15/06: Steve has received the xsds (there are three of them) and will review.<br>28/06: Not reviewed yet<br>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.<br>12/07: No further progress<br>19/07: Suman will talk to IBM legal and make it clear that the license must allow users to actively use the xsds.<br>26/07: No update<br>16/08: Steve has reviewed and sent comments to Suman. Suman has asked legal to clarify re-distribution rules, taking into account OGF licensing terms and patent filing.<br>23/08: No update |
| 133 | **Make a set of default formats available (Suman)**<br>19/01: Suman expects some default formats to be ready by Feb 9th. Will need approval to publish<br>26/01: Stephanie sent the defaults used by test cases to Suman<br>02/02: no update<br>09/02: no update<br>23/02: no update<br>02/03: Same as 132<br>09/03: No update. Same license issue apply though.<br>13/04: No update.<br>04/05: IBM will make one default format available. Suman is working through the IBM process to permit publication.<br>18/05: Awaiting response from IBM legal.<br>01/06:<br>08/06: Awaiting response from IBM legal. IBM also want to prove that the default format has the properties sensibly defined to plans to include in internal testing.<br>15/06: No change.<br>...<br>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).<br>26/07: No update<br>16/08: Same issue as action 132 |

| | |
|---|---|
| | 23/08: No update |
| 136 | **Arrays with missing elements (Steve)**<br>There is a problem when there are empty/missing array elements with an index greater than minOccurs. For example:<br><br>xs:element name"array" minOccurs=0, maxOccurs=10 lengthKind='delimited'<br><br>Datastream: ,,,value3,value4<br><br>Infoset will contain:<br>array[1] = value3<br>array[2] = value4<br><br>This is because elements with an index greater than minOccurs are optional and so do not get defaulted.<br><br>Unparsing this infoset will produce:<br><br>Datastream: value3,value4<br><br>You could make the empty space (%ES;) the nil value which will work for simple elements but not for complex.<br><br>Infoset will contain:<br>array[1] = nil<br>array[2] = nil<br>array[3] = value3<br>array[4] = value4<br>23/02: Discussed options<br>1. Changed definition of required for arrays to be 'required up to the last instance in the data stream of the array'<br>2. add index to the element info item<br>Steve to investigate if XDM uses an index.<br>02/03: No progress<br>09/03: No update<br>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.<br>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.<br>04/05: In progress.<br>18/05: In progress, at minimum some clarification to the spec is needed<br>01/06: Discussed to bring Mike up to speed. Solution will depend on other spec discussions.<br>08/06: Still under discussion<br>15/06: Will come back to this after action 140 resolved<br>28/06: As above<br>...<br>26/07: As above<br>16/08: See action 140<br>23/08: See action 140 |
| 140 | **Spec issue: Parsing: 'missing' v 'empty', role of initiators , default values (All)**<br>01/06: See minutes.<br>08/06: Still under discussion. Tim has sent Mike a selection of data formats to guide the discussion.<br>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.<br>05/07: Next extra call is Wed 6th July - Steve to send invite<br>12/07: Two more calls held. Next call is Wed 13th July.<br>19/07: More calls held, next call is Fri 22nd July.<br>26/07: More calls held, good progress<br>16/08: Steve will set up next call for when Tim has returned from holiday<br>23/08: Two more calls scheduled for this week, remaining issues: separator suppression, empty strings, sparse arrays (see action 136) |
| 141 | **Should text number exponent , infinity and Nan rep properties be lists ? (IBM)**<br>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?<br>05/07: In progress with IBM.<br>12/07: Still with IBM<br>19/07: Still with IBM<br>26/07: Still with IBM<br>16/08: Still with IBM<br>23/08: Still with IBM, a meeting will be held this week |
| 145 | **Provide a 'dispatch' way of discriminating a choice for better performance of the envelope/payload use case (Steve, Mike, Suman)**<br>12/7: See minutes. Need to choose a proposal and flesh out.<br>19/07: Waiting for proposals<br>26/07: Waiting for proposals<br>16/08: Waiting for proposals. Suman added to action.<br>23/08: Steve to send a proposal |
| 146 | **Model NRL's bit-oriented format where complex repeating element carries a 'last' indicator (Steve, Mike)**<br>19/07: NRL to send xsd and example of format.<br>26/07: Steve/Tim put forward a proposal that used a discriminator that looked at the previous item's repeatBit in the array and failed if it was 0. This looks like it will work. Mike to add the correct DFDL annotations to take into account both repeatBit and presentBit in a single discriminator, and auto-set them using outputValueCalc.<br>A more usable solution to the 'last indicator' problem is desirable though. Noted that this has a 'repeat..until' semantic.<br>16/08: Adam/Ryan to evaluate Steve/Tim proposal as it is potentially fragile relying as it does on backtracking.<br>23/08: Adam/Ryan had meeting with Boyd Fletcher and others, still evaluating to see whether change to spec is needed to handle this better. |
| 147 | **Clarify the rules when for padding /trimming and nil handling interact (Steve, Mike)**<br>19/07: 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. There is an inconsistency here, but is that ok?<br>26/07: No further thought on this<br>16/08: No further thought on this. Needs sub-team to evaluate.<br>23/08: No progress, will be discussed when current agenda items are completed |
| 148 | **Clarify the rules around pattern -based lengths and scanability (Mike)**<br>26/07: Mike to provide words<br>16/08: Mike and Steve have discussed via e-mail and will report back to WG when concluded.<br>23/08 Steve to write this up for next call |
| 150 | **What is missing from the DFDL infoset to enable XSD PSVI to be built from it ? (Steve)**<br>16/08: Issue raised by the XSD WG.<br>23/08: Two things missing from DFDL infoset for PSVI production, whether an element is valid, and, for a simple element with a union type, which member the value matched. Agreed that solving this makes the infoset more useful for XML oriented consumers.<br>Proposed:<br>1) Union member - fix by adding a new member SCD to infoset<br>2) Validity flag - fix by adding a new member to infoset items -> valid, invalid.<br>Both new members are empty if validation has not been enabled<br>This is fine for validating when parsing, but leaves a couple of questions: |

| | |
|---|---|
| | - What do these properties mean when unparsing?<br>- What happens if validation is enabled during unparsing? |
| 151 | **Improve DFDL Unicode BOM handling (Steve, Mike, Tim)**<br>23/08: Refine Steve's proposal for improved document level support for Unicode BOMs. |

## Closed actions

| No | Action |
|---|---|
| | |

## Deferred actions

| No | Action |
|---|---|
| 129 | **Press release to publicise DFDL (Steve)**<br>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.<br>17/11: no progress<br>...<br>08/12: Still no response from IBM press office<br>15/12: no progress<br>....<br>09/03: No progress<br>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)**<br>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.<br>15/12: Alan tested against test dfdl schema which worked correctly (after fixing some errors in the schema)<br>22/12: no update<br>12/01: Joe has some defects to fix before making available on gridforge.<br>19/01: There is a difficult problem to solve before Joe make the style sheet public<br>26/01: Working on problems<br>02/02: no progress<br>09/02: As it wasn't a simple as exoected this will be treated as a low priority action<br>23/02: Low prioity<br>09/03: Low priority<br>30/03: Deferring for now |
| 066 | **Investigate format for defining test cases (All)**<br>25/11:IBM to see if it is possible to publish its test case format.<br>04/12: no update<br>...<br>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'<br>24/03: No progress<br>03/03: Discussions have been taking place on the subset of tests that will be provided.<br>10/03: work is progressing<br>17/03: work is progressing<br>31/03: work is progressing<br>14/04: And XML test case format has been defined and is being tested.<br>21/04. Schema for TDML defined. Need to define how this and the test cases will be made public<br>05/05: Work still progressing<br>12/05: Work still progressing<br>02/06: Work still progressing on technical and legal considerations<br>...<br>25/08: Will chase to allow Daffodil access to test cases. The WG should define how implementation confirm that they 'conform to DFDL v1'<br>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.<br>08/09: IBM still progressing<br>15/09: IBM still progressing, expect tests to be available within a few weeks<br>22/09: IBM still progressing, expect tests to be available within a few weeks<br>29/09:Test cases are being prepared.<br>06/10: Some test cases should be available next week. Steve would like to be able to show the test case information at OGF 30.<br>13/10: Still progressing<br>10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally ones that fit the 'extended conformance' of NCSA Daffodil<br>17/11: Work is progressing on verifying the test cases. It should be possible to distribute to the WG in 2 weeks.<br>24/11: About half the test cases have been completed and are being reviewed internally.<br>01/12: Test cases should be available shortly<br>08/12: The test cases are in internal IBM review. Probably need a bit of reorganising before publication<br>Stephanie gave a brief overview of the format of the test cases.<br>15/12: Ruth joined the call to provide the latest status. The test cases have been updated and a draft read.me produced. Although not ready for public distribution Ruth will send them to Joe for feedback.<br>22/12: Test cases were sent to Joe for initial testing which found some problems in the Daffodil parser<br>12/01: All current tests use a default format whih Daffodil doesn't currently support. Joe suggested that there should be test that defined the same function using different definition forms. Also suggested that default formats should be provided by the WG. This had always been the intention. Action 133 raised to track.<br>19/01: There is currently no resource available in IBM to make more tests available. IBM to discuss how/if it can make a 'minimal compliance test suite' available.<br>26/01: Action kicked off within IBM. There was a brief discussion abot naming and organisation of test cases but no preferences were expressed<br>02/02: IBM will not have the resources to develop a full test suite in the near future. Steve suggested that we produce a list of required test cases so that anyone could supply them.<br>09/02: Steve had previously sent a list of areas to be tested. Please review.<br>23/02: Please review Steve's list of areas to be tested<br>02/03: Alan had reviewed Steve's list and we went through his comments. Agreed there is no need for separate tests for the infoset or for dfdl: property lists, unions etc but comment will be added that these should be exercised during property testing.<br>09/03: Alan updated the test document. Need more introduction and perhaps adopting the OGF template.<br>30/03. Ownership of test document passed to Steve. This action is merged with 112 and will cover all aspects of compliance suite.<br>13/04: IBM will not have time to create a compliance suite in the near future. Probably best to make this action deferred for now. |

## Work items:

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