IBM TeamRo

TeamRoom Plus

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

Meeting about Meetings\OGF

Project DFDL 1.0

Meeting Date 23-Aug-11 (Tues)
Meeting Time 15:00 - 16:00

Created by Steve Hanson on 09-Mar-11 Last Modified by Steve Hanson on 24-Aug-11

OGF DFDL Working Group Call, 23 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. 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. Last remaining issue:

12.3.7.1.3. Byte order mark

Byte-order marks are explicitly stated to be "not characters" in the Unicode standard.

No such thing as a BOM codepoint in a UTF-8 string. A UTF-8 byte sequence might encode the character code for a BOM, but this would be a meaningless inclusion of a BOM character code in a context where it will never be interpreted. I suggest that we drop the term UTF-8 here, and BOM's that get encoded when they are interpreted as character codes, and translated by the UTF-8 encoding algorithm into a multi-byte UTF-8 byte sequence, is handled the same way as other non-characters, i.e., what do we do when a high or low surrogate codepoint is present and we're to encode as UTF-8. I think the answer is we run the UTF-8 encode/decode algorithm, and whatever Unicode character code it creates is what it creates, and if that happens to come out as any of the non-characters (BOMs, surrogates, others perhaps), so be it.

The topic is about Unicode non-characters, not specifically BOMs.

The general topic is encoding/decoding our infoset Unicode character codes which have no real representation in the specified encoding.

2. 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 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).

3. 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?

4. 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) Mike Beckerle (Deloitte) Tim Kimber (IBM) Ryan Farrell (US NRL) Adam Fox (US NRL)

Apologies

Suman Kalia (IBM)

1. 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. Last remaining issue:

12.3.7.1.3. Byte order mark

Discussed the e-mail thread on BOMs between Steve, Mike and Tim.

Stable positions are either that BOM is just another character (as spec today) or that DFDL conforms to the Unicode specification. Agreed that the former does not fit the 'makes the simple things simple' goal and that DFDL modelers using Unicode wil expect BOMs to be handled. This is certainly true of IBM's customers. Agreed that document level BOM processing is needed, not element level - this is in keeping with Unicode spec. Started to review Steve's proposal in the e-mail but ran out of time. New action 151 raised to complete.

Meeting closed

16:05 UK

Next call

Tues 30th Aug 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
	Improve DFDL Unicode BOM handling (Steve, Mike, Tim)
	23/08: Refine Steve's proposal for improved document level support for Unicode BOMs.

Current Actions:

No	Action					
123						
	13/10: Draft of first 3 chapters has been written and will be distributed to WG					
	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!					
	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.					
	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					
	01/12: Alan has started lesson 4 which covers fixed and variable fields and arrays. 08/12: Alan has almost completed lesson 4. Will send out for review.					
	15/12: First draft of lesson 4 is available for review. Alan to send to Bob and Joe.					
	22/12: Alan has distributed drafts for tutorials on Basic Structure and Optional/Repeating elements. Please review					
	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.					
	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.					

26/01: No comments received about 3 tuorials distributed last week. Alan is still working on Text representation.

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.

09/02: Steve had reviewed tutorials 3,4,5 and updated versions have been distributed. Joe reviewed lesson on text elements.

Main points. Using 'represented as text' is confusing. Examples are too cluttered. Suggest simple targeted examples but still build up to final complete schema

23/02: New versions distributed and Steve has commented.

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.

09/03: Alan distributed an update to the text tutorial. Please review.

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.

13/04: Steve is working on the text properties tutorial.

04/05: No progress

18/05: No progress

01/06: No progress

08/06: No progress

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.

28/06: On hold.

...

26/07: On hold 16/08: On hold 23/08: On hold

124 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 when 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 tutorial 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 an 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 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.

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 examples.

19/07: Work has started to address the two issues noted on 12/07.

26/07. Changes made, Steve will distribute to the WG prior to posting on the OGF web site and updating the Wikipedia link.

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)

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 effect use 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.

26/07: No update

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.

23/08: No update

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 has 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).

26/07: No update

16/08: Same issue as action 132

23/08: No update

136 Arrays with missing elements (Steve)

There is a problem when there are empty/missing array elements with an index greater than 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'

2. add index to the element info item

Steve to investigate if XDM uses an index.

02/03: No progress 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

26/07: As above 16/08: See action 140 23/08: See action 140

Spec issue: Parsing: 'missing' v 'empty', role of initiators, default values (All)

140 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.

19/07: More calls held, next call is Fri 22nd July.

26/07: More calls held, good progress

16/08: Steve will set up next call for when Tim has returned from holiday

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)

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

26/07: Still with IBM

16/08: Still with IBM

23/08: Still with IBM, a meeting will be held this week

Provide a 'dispatch' way of discriminating a choice for better performance of the envelope/payload use case (Steve, Mike, Suman)

12/7: See minutes. Need to choose a proposal and flesh out.

19/07: Waiting for proposals

26/07: Waiting for proposals

16/08: Waiting for proposals. Suman added to action.

23/08: Steve to send a proposal

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.

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.

A more usable solution to the 'last indicator' problem is desirable though. Noted that this has a 'repeat..until' semantic. 16/08: Adam/Ryan to evaluate Steve/Tim proposal as it is potentially fragile relying as it does on backtracking. 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. 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 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? 26/07: No further thought on this 16/08: No further thought on this. Needs sub-team to evaluate. 23/08: No progress, will be discussed when current agenda items are completed 148 Clarify the rules around pattern -based lengths and scanability (Mike) 26/07: Mike to provide words 16/08: Mike and Steve have discussed via e-mail and will report back to WG when concluded. 23/08 Steve to write this up for next call What is missing from the DFDL infoset to enable XSD PSVI to be built from it ? (Steve) 150 16/08: Issue raised by the XSD WG. 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. Proposed: 1) Union member - fix by adding a new member SCD to infoset 2) Validity flag - fix by adding a new member to infoset items -> valid, invalid. Both new members are empty if validation has not been enabled This is fine for validating when parsing, but leaves a couple of questions: - What do these properties mean when unparsing? - What happens if validation is enabled during unparsing?

Closed actions

No	Action					
149						
	16/08: Do we support IEEE 754-1985 (as per XSDL 1.0) or -2008 (as per XSDL 1.1) or b					
	Do we allow physical lengths of 2 or 16?					
	23/08: Closed. Steve spoke with XML Schema WG. They moved to -2008 for 1.1 only					
because of new decimal support. That's why there are still just the xs:float and xs						
	types.					
Further, any support for these would very likely be by adding new built-in types						
from xs:anySimpleType, because precision differences mean the lexical spaces						
subsets.						
Decided that for DFDL:						
	- xs:float must have a physical length of 4 bytes for both 'ieee' and 'ibm390Hex' (schema					
	definition error otherwise).					
	- xs:double must have a physical length of 8 bytes for both 'ieee' and 'ibm390Hex' (schema					
	definition error otherwise).					
	- Add statement that if 'ibm390Hex' then there may be precision/rounding issues when					
	converting to/from infoset float which is IEEE					
	- Half-precision IEEE and quad-precision IEEE/IBM not supported					

Deferred actions

	No	Action						
ſ	129	9 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

12/01: Joe has some defects to fix before making available on gridforge.

19/01: There is a difficult problem to solve before Joe make the style sheet public

26/01: Working on problems

02/02: no progress

09/02: As it wasn't a simple as exoected this will be treated as a low priority action

23/02: Low prioity 09/03: Low priority

30/03: Deferring for now

066 Investigate format for defining test cases (All)

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

31/03: work is progressing

14/04: And XML test case format has been defined and is being tested.

21/04. Schema for TDML defined. Need to define how this and the test cases will be made public

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

15/09: IBM still progressing, expect tests to be available within a few weeks

22/09: IBM still progressing, expect tests to be available within a few weeks

29/09:Test cases are being prepared.

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.

13/10: Still progressing

10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally ones that fit the 'extended conformance' of NCSA Daffodil

17/11: Work is progressing on verifying the test cases. It should be possible to distribute to the WG in 2 weeks.

24/11: About half the test cases have been completed and are being reviewed internally.

01/12: Test cases should be available shortly

08/12: The test cases are in internal IBM review. Probably need a bit of reorganising before publication

Stephanie gave a brief overview of the format of the test cases.

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.

22/12: Test cases were sent to Joe for initial testing which found some problems in the Daffodil parser

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.

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.

26/01: Action kicked off within IBM. There was a brief discussion abot naming and organisation of test cases but no preferences were expressed

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.

09/02: Steve had previously sent a list of areas to be tested. Please review.

23/02: Please review Steve's list of areas to be tested

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.

09/03: Alan updated the test document. Need more introduction and perhaps adopting the OGF template.

30/03. Ownership of test document passed to Steve. This action is merged with 112 and will cover all aspects of compliance suite.

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.	Steve	N/A	Ongoing
	http://forge.gridforum.org/sf/go/doc16280?nav=1			
044	Incorporate errata list into DFDL spec.	Steve	TBD	