

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	10-Jan-12 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 10-Jan-12

OGF DFDL Working Group Call , 10 January 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. Inconsistent naming of text bidi properties

See e-mail - the names could be more consistent

2. Property documentFinalTerminatorCanBeMissing not flexible enough

See e-mail - unable to handle the case where final terminator is tolerated when parsing but must not be output. Does DFDL 1.0 need to handle this?

3. UTF-8 in DFDL really does mean UTF -8 and not CESU-8 etc

See e-mail - needs to be clarified.

4. ISO 10646 on its own not sufficient to represent all characters ?

See e-mail - Mike is proposing some updated words to cover this.

5. Possible change of time for call .

Can attendees make 1 hour later?

6.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) (co-chair)

Mike Beckerle (co-chair)

Steph Fetzer (IBM)

Kristoffer Rose (IBM)

Jonathan Cranford (Mitre Corp)

Alonso Mumford (US NSA)

Apologies

Tim Kimber (IBM)

Minutes

1. Inconsistent naming of text bidi properties

Agreed to change names to remove any second occurrence of 'text' from bidi properties.
Errata taken.

2. Property documentFinalTerminatorCanBeMissing not flexible enough

DFDL is unable to model the case where final terminator is tolerated when parsing but must not be output when unparsing. **Action 159** raised to explore whether this capability is really needed, or whether an alternative exists (eg, use separator).

3. UTF-8 in DFDL really does mean UTF -8 and not CESU-8 etc

Agreed to clarify this. Errata taken.

4. ISO 10646 on its own not sufficient to represent all characters ?

Mike is proposing some updated words to cover this. He will revise and submit.
Action 160 raised to cover this.

5. Possible change of time for call .

All subsequent DFDL WG calls will start an hour later at 4pm UK / 11am Eastern.

6. Expressions and empty string

Agreed that properties that use empty string to indicate that the property is not applicable, and can also take an expression as the value, should not allow the expression to return empty string.

7. Daffodil Open Source Project

Mike reported on the latest status of the project, which is ramping up.
US NSA and DoD labs plus at least three contracting firms are participating.
Mailing list started with 16 people currently subscribed.
Aim is to have working parser by end 2012, unparsing working by end 2013
These are fully conformant implementations, ie, all optional features supported.
It would be useful to demonstrate a milestone in 2012 with a parser that implements a minimal subset, eg, fixed length binary data including bits. That would then enable the WG to present a DFDL Update session at a future OGF meeting that covered both the IBM and Daffodil implementations, and which was backed by downloadable Daffodil code.

Meeting closed

15:45 UK

Next call

Tues 17th Jan 2012 @ 16:00 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: 161

Actions raised at this meeting

No	Action
159	Property documentFinalTerminatorCanBeMissing not flexible enough ? (Steve) 10/01: Do we need to upgrade this property to handle formats where the terminator is not allowed on output but tolerated on input?
160	ISO Errata: 10646 on its own not sufficient to represent all characters ? (Mike) 10/01: Analysis has shown this to be the case. Mike will supply words for the errata.

Current Actions :

No	Action
123	DFDL tutorial (Steve) 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.

	<p>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.</p> <p>26/01: No comments received about 3 tuorials distributed last week. Alan is still working on Text representation.</p> <p>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.</p> <p>09/02: Steve had reviewed tutorials 3,4,5 and updated versions have been distributed. Joe reviewed lesson on text elements.</p> <p>Main points. Using 'represented as text' is confusing. Examples are too cluttered. Suggest simple targeted examples but still build up to final complete schema</p> <p>23/02: New versions distributed and Steve has commented.</p> <p>02/03: Alan has published the final versions of tutorials 4,5,6 and is working on text representations. 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.</p> <p>09/03: Alan distributed an update to the text tutorial. Please review.</p> <p>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.</p> <p>13/04: Steve is working on the text properties tutorial.</p> <p>04/05: No progress</p> <p>18/05: No progress</p> <p>01/06: No progress</p> <p>08/06: No progress</p> <p>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.</p> <p>28/06: On hold.</p> <p>...</p> <p>29/11: Tim offered to take a look at the next outstanding tutorials. Steve / Tim to discuss</p> <p>6/12: No progress</p> <p>...</p> <p>10/01: No progress, offer from Mike to help. First step is to make any corrections due to errata.</p>
133	<p>Make a set of default formats available (Suman)</p> <p>19/01: Suman expects some default formats to be ready by Feb 9th. Will need approval to publish</p> <p>26/01: Stephanie sent the defaults used by test cases to Suman</p> <p>02/02: no update</p> <p>09/02: no update</p> <p>23/02: no update</p> <p>02/03: Same as 132</p> <p>09/03: No update. Same license issue apply though.</p> <p>13/04: No update.</p> <p>04/05: IBM will make one default format available. Suman is working through the IBM process to permit publication.</p> <p>18/05: Awaiting response from IBM legal.</p> <p>01/06:</p> <p>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 .</p> <p>15/06: No change.</p> <p>...</p> <p>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).</p> <p>26/07: No update</p> <p>16/08: Same issue as action 132</p> <p>23/08: No update</p> <p>30/08: As action 132</p> <p>20/09: Steve has mailed IBM Hursley legal</p> <p>27/09: Steve to meet with IBM Hursley legal</p>

	<p>4/10: No further progress</p> <p>18/10: Steve has met with legal, can publish the xsds after tidy up. Will create OGF info document.</p> <p>1/11: No further progress.</p> <p>...</p> <p>29/11: Steve has a version of the default format available, waiting for one last IBM legal procedure.</p> <p>6/12: Waiting for one last IBM legal procedure.</p> <p>13/12: Still waiting for one last IBM legal procedure. OK to review iwhat we have so far though, so Steve sent out an xsd containing a single defineFormat and a single defineEscapeScheme.</p> <p>The spec in section 19 makes it clear that any such xsds are essentially examples, so this xsd will be treated as such. It will be in a 'DFDL example' namespace and we will publish as OGF information document on GridForge.</p> <p>Reviewed the properties and changed occursCountKind to 'parsed'. (Noted that 'implicit' is better but that is subject of an errata not yet folded back into the spec).</p> <p>Stayed with lengthKind 'delimited' with a comma separator so that the xsd would actually model a real single record CSV stream.</p> <p>Steve to improve the example usage section in the prolog to include namespaces and example data, and run some real CSV tests.</p> <p>10/01: IBM legal process complete. Steve has updated the xsd as above and used it to model CSV data. Steve to formalise as an OGF document.</p>
140	<p>Spec issue: Parsing: 'missing' v 'empty', role of initiators, default values (All)</p> <p>01/06: See minutes.</p> <p>08/06: Still under discussion. Tim has sent Mike a selection of data formats to guide the discussion.</p> <p>15/06: Not discussed - an extra call has been scheduled to go through this.</p> <p>28/06: A series of extra calls are being held between Mike, Steve, Tim and Steph.</p> <p>05/07: Next extra call is Wed 6th July - Steve to send invite</p> <p>12/07: Two more calls held. Next call is Wed 13th July.</p> <p>19/07: More calls held, next call is Fri 22nd July.</p> <p>26/07: More calls held, good progress</p> <p>16/08: Steve will set up next call for when Tim has returned from holiday</p> <p>23/08: Two more calls scheduled for this week, remaining issues: separator suppression, empty strings, sparse arrays (see action 136)</p> <p>30/08: Call held earlier today. Still remaining - separator suppression (matrix); sparse arrays; empty strings; empty value delimiter policy. Steve to summarise where we have got so far before remaining items are tackled.</p> <p>20/09: Steve has summarised where we are with action 140, which Tim and Mike have reviewed. 2 hours call planned for Thursday.</p> <p>27/09: Calls held, progressing the separator suppression behaviour</p> <p>4/10: More calls held, progressing separator suppression, sparse arrays (see action 136) and emptyValueDelimiterPolicy behaviour</p> <p>18/10: All issues now in a single document, call held earlier today. Next call Thursday.</p> <p>1/11: Extra calls still ongoing</p> <p>8/11: Extra calls still going, when action 140 document complete will send to Steph for review</p> <p>15/11: Mike is verifying the action 140 conclusions by writing an algorithm in SCALA which can be ultimately be used in Daffodil</p> <p>22/11: Call to be held next week</p> <p>29/11: Next call 30/11</p> <p>6/12: Next call 7/12</p> <p>13/12: Had call earlier today, making good progress. Next call first week of Jan.</p> <p>10/01: Next call to be scheduled for Wed or Fri</p>
145	<p>Provide a 'dispatch' way of discriminating a choice for better performance of the envelope/payload use case (Steve, Mike, Suman)</p> <p>12/7: See minutes. Need to choose a proposal and flesh out.</p> <p>19/07: Waiting for proposals</p> <p>26/07: Waiting for proposals</p> <p>16/08: Waiting for proposals. Suman added to action.</p> <p>...</p>

	<p>1/11: Steve to send a proposal</p> <p>...</p> <p>10/01: Waiting for Steve to find time to do this</p>
146	<p>Model NRL's bit-oriented format where complex repeating element carries a 'last' indicator (Steve, Mike)</p> <p>19/07: NRL to send xsd and example of format.</p> <p>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.</p> <p>A more usable solution to the 'last indicator' problem is desirable though. Noted that this has a 'repeat..until' semantic.</p> <p>16/08: Adam and Ryan to evaluate Steve/Tim proposal as it is potentially fragile relying as it does on backtracking.</p> <p>23/08: Adam and Ryan had meeting with Boyd Fletcher and others, still evaluating to see whether change to spec is needed to handle this better.</p> <p>30/08: No update</p> <p>20/09: Looking like a spec change would be beneficial. Adam will provide details within the next month.</p> <p>27/09: Waiting to hear from NRL. Mike to find out how they are progressing.</p> <p>4/10: Mike will contact NRL to see how things are progressing.</p> <p>18/10: No update - Steve & Mike to chase</p> <p>1/11: No update</p> <p>8/11: Update on NRL via Mike - staffing the team to carry forward the DFDL work</p> <p>15/11: No update</p> <p>22/11: No update</p> <p>29/11: NSA still expressing an interest in Daffodil</p> <p>6/12: No update</p> <p>13/12: NSA held a meeting with Steve & Mike to discuss their contributing to Daffodil through 2012</p> <p>10/01: Will be added as an action item for the Daffodil Open Source Project</p>
153	<p>Decide which TZ scheme to adopt to extend dfdl :calendarTimeZone (Steve)</p> <p>18/10: Need to keep as close to ICU as possible as an aid to implementers.</p> <p>1/11: No progress</p> <p>...</p> <p>13/12: Waiting for Steve to find time to do this</p> <p>10/01: Steve looking at this now, will have a proposal for next week</p>

Closed actions

No	Action
132	<p>Publishing DFDL xsd (Suman)</p> <p>08/12: Agreed that it should be made available. Suman has started the approval process in IBM</p> <p>15/12: no progress</p> <p>22/12: no update</p> <p>12/01: Suman is getting approval from IBM to publish.</p> <p>19/01: Waiting to get IBM approval to publish</p> <p>26/01: no update</p> <p>02/02: No update</p> <p>09/02: no update</p> <p>23/02: no update</p> <p>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</p> <p>09/03: No update</p> <p>30/03: Suman has sent information to IBM legal. Reminded him about the license issue.</p> <p>13/04: No update.</p> <p>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</p>

	<p>needed.</p> <p>18/05: Awaiting response from IBM legal. Suman will send Steve the model xsd for review.</p> <p>01/06:</p> <p>08/06: Awaiting response from IBM legal. Suman will send Steve the model xsd for review.</p> <p>15/06: Steve has received the xsds (there are three of them) and will review.</p> <p>28/06: Not reviewed yet</p> <p>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.</p> <p>12/07: No further progress</p> <p>19/07: Suman will talk to IBM legal and make it clear that the license must allow users to actively use the xsds.</p> <p>26/07: No update</p> <p>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.</p> <p>23/08: No update</p> <p>30/08: Steve has reviewed words from IBM legal. It looks like that once the xsds are published by OGF then they are free to be used under the normal OGF licensing terms. Steve needs to clear all this with IBM legal in Hursley, as the DFDL mission is owned by Hursley in UK.</p> <p>20/09: Steve has mailed IBM Hursley legal</p> <p>27/09: Steve to meet with IBM Hursley legal</p> <p>4/10: No further progress</p> <p>18/10: Steve has met with legal, can publish the xsds after tidy up. Will create OGF info document.</p> <p>1/11: No further progress</p> <p>...</p> <p>29/11: Waiting for Steve to find time to do this</p> <p>6/12: Waiting for one last IBM legal procedure.</p> <p>13/12: Still waiting for one last IBM legal procedure.</p> <p>10/01: Closed. IBM legal process complete. Steve has sent xsds to Mike who has reviewed and spotted a couple of errors which Steve/Suman have fixed.</p>
156	<p>Errata: Investigate ICU fallback mappings when converting from Unicode (Mike)</p> <p>8/11: Need more information before errata can be taken</p> <p>15/11: Action assigned to Mike</p> <p>22/11: Mike making good progress - will send update by email. Unicode Character Mapping Markup Language might be able to help here.</p> <p>29/11: Mike has sent a detailed V001 proposal via email, Steve will comment and respond.</p> <p>6/12: Steve has reviewed and sent response, as has Tim. Andreas Martens from IBM joined the call as he has a lot of experience with ICU's encoding support and real-world requirements.</p> <p>Mike's proposal is workable but the feeling was that DFDL did not need the complexity of a new annotation type.</p> <p>It was suggested that the following was sufficient:</p> <ul style="list-style-type: none"> - Parsing: Allow modeler to choose whether to Skip, Error or Substitute when bad characters encountered. Substitute would use U+FFFD. - Unparsing: Allow modeler to choose whether to Skip, Error or Substitute when bad characters encountered. Substitute would use encoding's substitution character, modeler can override with this own, single, character. <p>Substitute behaves like Error if there is no substitution character defined.</p> <p>Note that ICU tries the encoding-defined primary mapping first, then any encoding-defined fallback mapping, then any substitution character.</p> <p>Decided not to include support for ReplaceWithHex as the DFDL modeler can always model the data as a BLOB if he wants everything preserved without loss/change.</p> <p>Mike will revise his proposal as V002 and redistribute.</p> <p>13/12: Mike's simplified proposal looks good, in summary there is one new (standard) property dfdl:encodingErrorPolicy with values 'skip', 'error', 'replace' that applies on both parsing and unparsing. There remain a few questions about invalid bytes/characters when parsing and what ICU is able to detect. Mike will experiment and report back.</p> <p>10/01: Closed. Mike has done some investigation and has proposed words to cover this. Errata will be taken. This also spawned two more agenda items.</p>

Deferred actions

No	Action
129	<p>Press release to publicise DFDL (Steve)</p> <p>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.</p> <p>17/11: no progress</p> <p>...</p> <p>08/12: Still no response from IBM press office</p> <p>15/12: no progress</p> <p>....</p> <p>09/03: No progress</p> <p>30/03: Making this action deferred until IBM is in a position to say something more concrete about any implementation.</p>
131	<p>Transformation of DFDL properties to a canonical form (Joe)</p> <p>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.</p> <p>15/12: Alan tested against test dfdl schema which worked correctly (after fixing some errors in the schema)</p> <p>22/12: no update</p> <p>12/01: Joe has some defects to fix before making available on gridforge.</p> <p>19/01: There is a difficult problem to solve before Joe make the style sheet public</p> <p>26/01: Working on problems</p> <p>02/02: no progress</p> <p>09/02: As it wasn't a simple as exoected this will be treated as a low priority action</p> <p>23/02: Low prioity</p> <p>09/03: Low priority</p> <p>30/03: Deferring for now</p>
066	<p>Investigate format for defining test cases (All)</p> <p>25/11: IBM to see if it is possible to publish its test case format.</p> <p>04/12: no update</p> <p>...</p> <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> <p>08/09: IBM still progressing</p> <p>15/09: IBM still progressing, expect tests to be available within a few weeks</p> <p>22/09: IBM still progressing, expect tests to be available within a few weeks</p> <p>29/09: Test cases are being prepared.</p> <p>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.</p> <p>13/10: Still progressing</p> <p>10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally</p>

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 which 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 about 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 info set 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. http://forge.gridforum.org/sf/go/doc16280?nav=1	Steve	N/A	Draft 0006 on grid forge.
044	Incorporate errata list into DFDL spec.	Steve	TBD	