TeamRoom Plus

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

Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	30-Mar-11 (Wed)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 31-Mar-11

OGF DFDL Working Group Call , 30 March 2011

Agenda

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

Minutes

Meeting Minutes

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

Open Grid Forum : Data Format Description Language Working Group

OGF DFDL Working Group Call , 30 March 2011

Attendees
Alan Powell (IBM)
Steve Hanson (IBM)
Bob McGrath (NCSA)
Stephanie Fetzer (IBM)

Apologies
Joe Futrelle (NCSA)
Suman Kalia (IBM)

1. Current Actions
Updated below

2. Binary Packed Calendar elements
The spec is surprisingly short of details of binary packed and bcd calendar elements

From Tim
The DFDL spec is somewhat light on details when it describes binaryCalendarRep ='packed' and 'bcd'. It's not obvious how an implementer should interpret or write the packed/bcd number. or what patterns are allowed. It may be that the current wording is not too far wrong. However, it is likely that we need the VDP to be applied to the date or time, based on the information on this page:

http://msdn.microsoft.com/en-us/library/ee267429%28v=bts.10%29.aspx
(7 digits being used for times implies HHMMSS.S)

Clarify what packed and BCD calendars mean.
- No need to use a separate VDP property. The only place where a decimal point can occur is for fractional seconds. This is detectable from the pattern at the boundary of 's' and 'S', ie sS
- Property calendarPatternKind = 'explicit' must be used with binary calendar representations, as the defaults for 'implicit' use non-numeric characters. Schema defn error otherwise.
- Property binaryCalendarRep should restate the rule from property calendarPattern.
- Examples to be provided.
Errata will be raised against the spec.

3. AOB
Agreed to make meetings bi-weekly from now on.

DFDL web pages have been mentioned by OGF as a good example for other standards to follow.

Alan retiring from IBM on 1st April. Steve formally expressed the WG's appreciation of the work Alan has done to get the DFDL specification etc into the position it is now in.

Meeting closed
15:40

Next call
Wednesday 6 April 2011 15:00 UK (10:00 ET)

Create Action Items

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

Action Items and Other Meeting Documents

View: ResultDocs

Next action: **138**

Actions raised at this meeting

No	Action

Current Actions :

No	Action
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.
111	**Daffodil DFDL parser (Bob, Joe)**
	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.
	We discussed the scenarios that motivated the development which was to extract data from various sources and transform into canonical formats.
	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.
	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.
	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 ahd responding indicating the extra functions he thought were essential.
	Since then Alejandro has implemented some of the missing functions, such as escape schemes, pre-defined variables, binary decimal numbers, etc, and will update his list.
	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
	29/09: not discussed
	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.
	13/10: Bob still progressing project funding and making Daffodil publically available.
	10/11: NCSA internal & 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.
	17/11: Joe has started coming up to speed with Daffodil. Bob is waiting for signoff from the university to open source the code.

	24/11: not discussed 01/12: Joe is becoming familiar with the code. Waiting for university to give permission to make available. 08/12: Disclosure process in progressing. Bob has permission to make the web page available so hopefully that will be soon. The US National Archive have an article on their website about DFDL. Unfortunately it is a little out of date. 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 http://daffodil.ncsa.uiuc.edu/. 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."
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. 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.

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.
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.
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.

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.

137	**Map DFDL infoset to XDM (Steve)** This is an old workitem 02/03: No progress 09/03: No update 30/03: No progress

Closed actions

No	Action
112	**DFDL certification process** 25/08: Discussed how to certify DFDL implementations. Alan to investigate if OGF have a defined process. 01/09: In progress, spec needs to state what conformance means, as part of this work 08/09: Discussed what needs to be said in the spec and agreed that details of a conformance test suite should be in another document. Alan to draft conformance section. 15/09: Alan had look at the conformance sections in XML and Schema specifications both of which indicate sections which must be implemented. None just say 'execute the test suite'. They talk in terms of conformance of document, schema and processors.. 22/09: no progress 22/09: Alan has added short Conformance and Optional Features sections to spec which was briefly discussed. Discussed naming for processors that don't implement optional features and those that implement all features. 06/10: Need to decide what/how test cases and certification process should occur 13/10: no progress ... 08/12: no progress 15/12: no progress 22/12: no progress 12/01: no progress 19/01: Agreed that a minimum compliance test suite needs to be provided. This could be a test suite and/or a web service that tested DFDL implementations 26/01: Waiting for test suite. 02/02: We need to document the intention of providing a test suite as the compliance process. 09/02: Alan will create a 'DFDL conformance' document that will state that a test suite will be is being developed. 23/02: No update 02/03: No progress 09/03: See action 066 30/03: Closed. Action merged with 066.

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 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

Work items:

No	Item	Owner	Target	Status
042	Mapping of the DFDL infoset to XDM	Steve	N/A	Not required for V1 specification