DFDL Escape Characters

### Author: Avery Bibeau

# Explanation

Escape characters are one of two methods for escaping separator characters in DFDL elements. Escaping is a useful capability for using alternative data representation that includes the separator character. Escape characters can be used to essentially ignore the following separator character in an element.

# Problem

This tutorial explores escape characters with alternative data representation in a list of names where each name can be represented as ‘First Last’ or ‘Last, First’

The following data set will be parsed: a list of names containing two names where each name is separated by a comma. Notice that the second name uses a comma (the separator character), this character needs to be escaped by an escape character.

names=John Doe,Doe\, Jane

# Solution

A defineEscapeScheme is required to create properties related to escaping. This definition exists outside of the default properties schema.

1. The dfdl:defineEscapeScheme is created in the annotation, immediately below the schema’s dfdl:format and is given the name "DefaultPropertiesEscapeScheme"

<xsd:annotation>

<xsd:appinfo source="http://www.ogf.org/dfdl/">

<dfdl:format ref="DefaultPropertiesFormat"/>

<dfdl:defineEscapeScheme name="DefaultPropertiesEscapeScheme">

<dfdl:escapeScheme …> </dfdl:escapeScheme>

</dfdl:defineEscapeScheme>

</xsd:appinfo>

</xsd:annotation>

1. The property escapeKind set to "escapeCharacter" is necessary for using escape characters.
2. Next, the character used to escape the subsequent character is defined with the escapeCharacter property; this will be set to "\"
3. Next, the escapeEscapeCharacter is used to define a character for escaping the escapeCharacter, this property is set to "\". While this property is not entirely necessary for this example, it is still a useful capability.

The resulting XML infoset looks like this:

<names>

<name>John Doe</name>

<name>Doe, Jane</name>

</names>

1. To further explore the escapeEscapeCharacter, we can use the following data set

names=John Doe,Doe\, Jane\\Joan

1. This data set demonstrates escaping the escapeCharacter character with the escapeEscapeCharacter. The resulting XML infoset looks like this:

<names>

<name>John Doe</name>

<name>Doe, Jane\Joan</name>

</names>

# Schema

Here’s a complete schema using escape characters:

<?xml version="1.0" encoding="UTF-8"?>

<xsd:schema xmlns:dfdl="http://www.ogf.org/dfdl/dfdl-1.0/" xmlns:xsd="http://www.w3.org/2001/XMLSchema"

xmlns:fn="http://www.w3.org/2005/xpath-functions">

<xsd:include schemaLocation="DefaultProperties.xsd"/>

<xsd:annotation>

<xsd:appinfo source="http://www.ogf.org/dfdl/">

<dfdl:format ref="DefaultPropertiesFormat"/>

<dfdl:defineEscapeScheme name="DefaultPropertiesEscapeScheme">

<dfdl:escapeScheme escapeKind="escapeCharacter" escapeCharacter="\"escapeEscapeCharacter="\">

</dfdl:escapeScheme>

</dfdl:defineEscapeScheme>

</xsd:appinfo>

</xsd:annotation>

<xsd:element name="names" dfdl:initiator="names=">

<xsd:complexType>

<xsd:sequence dfdl:separator=",">

<xsd:element name="name" type="xsd:string" dfdl:occursCountKind="implicit" minOccurs="1" maxOccurs="unbounded" dfdl:escapeSchemeRef="DefaultPropertiesEscapeScheme"/>

</xsd:sequence>

</xsd:complexType>

</xsd:element>

</xsd:schema>