Megan Masanz

CS 598 – Data Curation Final Project

Contents

[Overview 1](#_Toc531546690)

[Terminology 2](#_Toc531546691)

[Part 1 Profile and MD5 checksum 2](#_Toc531546692)

[Part 1.1 2](#_Toc531546693)

[Part 1.2 Create a DTD for each file 3](#_Toc531546694)

[1.2.1 File A XML Document DTD Implementation 3](#_Toc531546695)

[1.2.1 .1 XML Prose for element, attribute, and attribute values for file A 7](#_Toc531546696)

[1.2.1 .2 Difficult decisions in designing DTD for file A XML 8](#_Toc531546697)

[1.2.2 File B XML Document DTD Implementation 9](#_Toc531546698)

[1.2.2.1 XML Prose for element, attribute, and attribute values for file B 13](#_Toc531546699)

[1.2.2.2 Difficult decisions in designing DTD for file B XML 14](#_Toc531546700)

[Part 1.3 Canonicalize the two data files and run checksums again to check for equivalence 15](#_Toc531546701)

[Part 1.4 DTD of Final XML File 17](#_Toc531546702)

[1.4.2.1 XML Prose for element, attribute, and attribute values for final generated file 29](#_Toc531546703)

[1.4.1 .2 Difficult decisions in designing DTD for file final XML 30](#_Toc531546704)

[Appendix 30](#_Toc531546705)

[*Figure Appendix.1.1 Code snipped to compute md5 hash string* 30](#_Toc531546706)

[*Figure Appendix.1.2.1 File A Embedded DTD Validation* 31](#_Toc531546707)

[*Figure Appendix.1.2.2 File B Embedded DTD Validation* 32](#_Toc531546708)

[*Figure Appendix 1.3.1 Winmerge File Comparison* 33](#_Toc531546709)

[*Figure Appendix 1.4.1 Embedded DTD Validation* 34](#_Toc531546710)

[References 34](#_Toc531546711)

# Overview

This project contains the output of an assessment of the data transfer quality of complaints information ensuring the equivalence of data in an old system and that in the new system. The information found below details the process of providing a solution to provide that analysis. In addition, this project contains an email notification remarking on the importance of data curation to the data science organization to be sent to its director.

To support the curatorial activity of preservation, this project is stored on github at: <https://github.com/megado123/cs598-finalproject>.

## Terminology

In this report there are two files referenced for comparison. Consumser\_Complaints\_FileA.xml will be referred to as File A. Consumer\_Complaints\_FileB.xml will be referred to as File B.

# Part 1 Profile and MD5 checksum

## Part 1.1

Consumser\_Complaints\_FileA.xml is a UT-F8 encoded XML file used by a government agency that tracks customer complaints against financial institutions. Consumser\_Complaints\_FileA.xml was generated by the legacy tracking system. The file does not contain DTD or XSD definitions. Line Feeds are used to format the file consistently and can be found in the unsorted data elements. The file utilizes attributes to hold information including identifiers, dates when the complaint was received and sent, as well as if the complaint was disputed and handled timely. The file contains elements which contain product information, company information, response, issue, and events. MD5 checksum: 637737835B3639596BF6DB0FA0FFF691

Consumer\_Complaints\_FileB.xml file is UT-F8 encoded XML file used by a government agency that tracks customer complaints against financial institutions. Consumer\_Complaints\_FileB.xml was generated by the current system. Line Feeds do not systematically separate unsorted elements, and comments can be found within the XML file. The XML file contains a minimal DTD consisting of an entity definition. The file utilizes attributes holding information including identifiers, submission type, dates when the complaint was received and sent, as well as if the complaint was disputed and handled timely. The file holds elements containing product information, company information, response, issue, and events. MD5 checksum: C2FB08E9A52DC8CD4D7B0C195061C783

Note: checksum values were populated through c# code commented and provided in [Appendix Figure Appendix.1.1](#_Figure_Appendix.1.1_Code) and available through github repository. For further validation <http://onlinemd5.com/> was used to confirm coded values as shown below in screenshots for Figure 1.1.A and Figure 1.1.B.

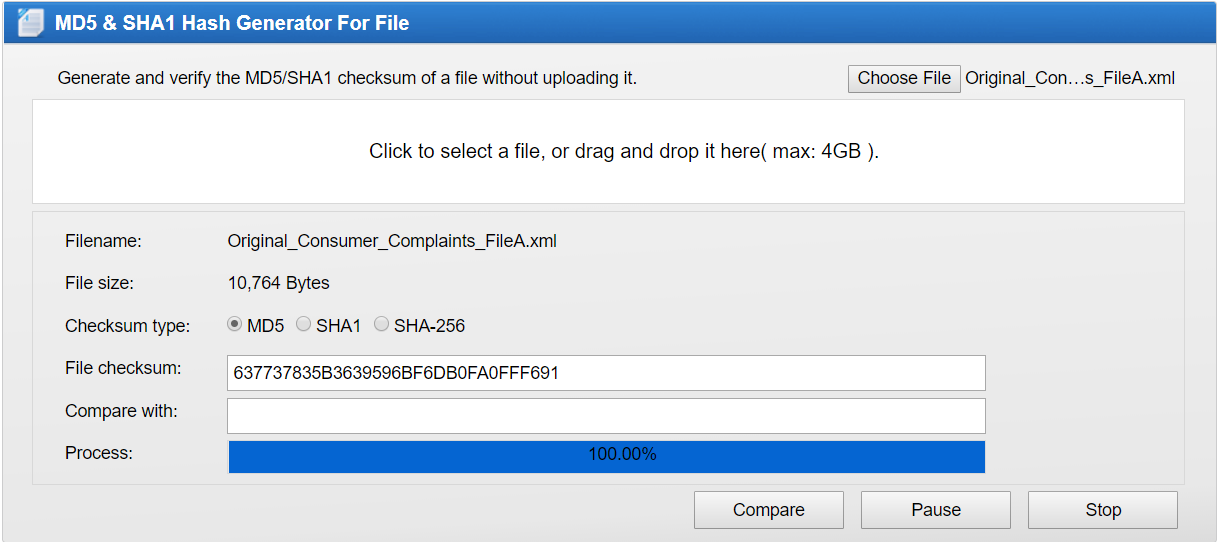


Figure 1.1.A – CheckSum validation from <http://onlinemd5.com/>

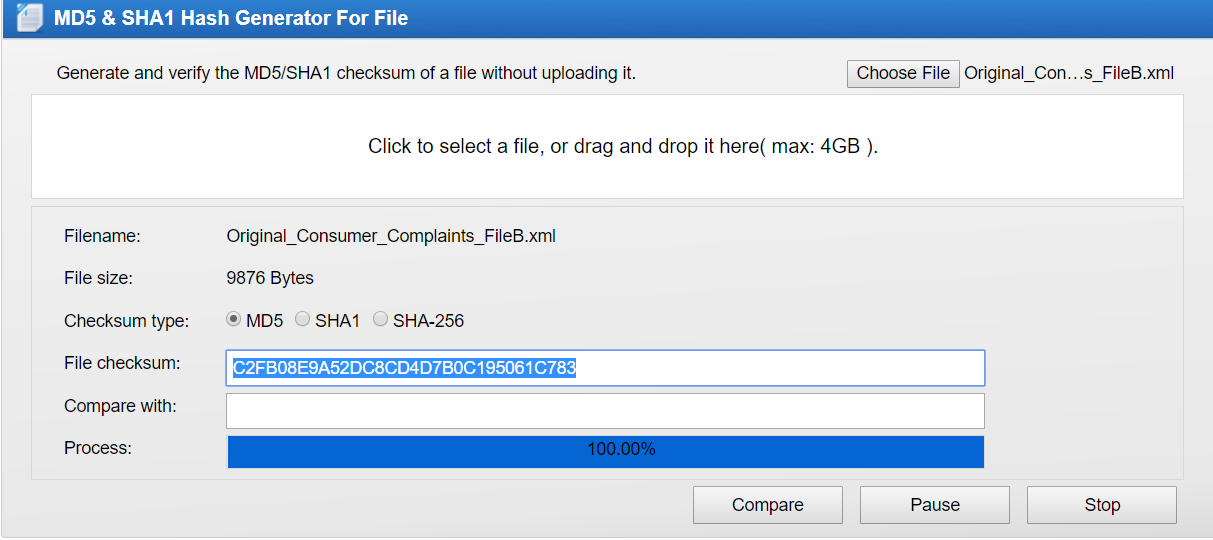


Figure 1.1.B – CheckSum validation from <http://onlinemd5.com/>

## Part 1.2 Create a DTD for each file

### 1.2.1 File A XML Document DTD Implementation

The file A with an embedded DTD is included in this solution folder structure under:

[\File A with Embedded DTD\FileA\_Consumer\_Complaints\_Embedded.xml](File_A_with_Embedded_DTD/FileA_Consumer_Complaints_Embedded.xml)

The file A with an external DTD is included in this solution folder structure under:

[\File\_A\_with\_External\_DTD\Consumer\_Complaints\_FileA\_with\_External\_DTD.xml](File_A_with_External_DTD/Consumer_Complaints_FileA_with_External_DTD.xml)

Below is the embedded DTD implementation for file A. The file has been validated using the website: <http://xmlvalidator.new-studio.org/> with a screen shot of its validity found in [Figure Appendix.1.2.1 File A Embedded DTD Validation](#_Figure_Appendix.1.2.1_File). This website requires embedded DTD’s but an external DTD has also been provided for completeness which was validated using the XMLSpy software package.

|  |
| --- |
| <?xml version="1.0"?>  <!DOCTYPE consumerComplaints [  <!ELEMENT consumerComplaints (complaint+)>  <!ELEMENT complaint ((event, event, product, issue, consumerNarrative?, company, submitted,response) | (company, event, submitted, issue, product, event, response))>  <!ELEMENT event EMPTY>  <!ELEMENT product (productType,subproduct?)>  <!ELEMENT productType (#PCDATA)>  <!ELEMENT subproduct (#PCDATA)>  <!ELEMENT issue (issueType, subissue?)>  <!ELEMENT issueType (#PCDATA)>  <!ELEMENT subissue (#PCDATA)>  <!ELEMENT consumerNarrative (#PCDATA)>  <!ELEMENT company (companyName,companyState,companyZip)>  <!ELEMENT companyName (#PCDATA)>  <!ELEMENT companyState (#PCDATA)>  <!ELEMENT companyZip (#PCDATA)>  <!ELEMENT submitted EMPTY>  <!ELEMENT response (publicResponse?, responseType)>  <!ELEMENT responseType (#PCDATA)>  <!ELEMENT publicResponse (#PCDATA)>  <!ATTLIST complaint id NMTOKEN #REQUIRED>  <!ATTLIST event type (received | sentToCompany) #REQUIRED >  <!ATTLIST event date CDATA #REQUIRED>  <!ATTLIST submitted via (Phone|Referral|Web) #REQUIRED>  <!ATTLIST response timely (Y|N) #REQUIRED>  <!ATTLIST response consumerDisputed (Y|N) #REQUIRED>  ]>  <consumerComplaints>  <complaint id="759222">  <event type="received" date="2014-03-12"/>  <event type="sentToCompany" date="2014-03-17"/>  <product>  <productType>Mortgage</productType>  <subproduct>Other mortgage</subproduct>  </product>  <issue>  <issueType>Loan modification,collection,foreclosure</issueType>  </issue>  <company>  <companyName>M&amp;T Bank Corporation</companyName>  <companyState>MI</companyState>  <companyZip>48382</companyZip>  </company>  <submitted via="Referral"/>  <response timely="Y" consumerDisputed="Y">  <responseType>Closed with explanation</responseType>  </response>  </complaint>  <complaint id="596562">  <event type="received" date="2013-11-13"/>  <event type="sentToCompany" date="2013-11-20"/>  <product>  <productType>Mortgage</productType>  <subproduct>Conventional adjustable mortgage</subproduct>  </product>  <issue>  <issueType>Loan servicing, payments, escrow account</issueType>  </issue>  <company>  <companyName>U.S. BANCORP</companyName>  <companyState>MN</companyState>  <companyZip>48322</companyZip>  </company>  <submitted via="Phone"/>  <response timely="Y" consumerDisputed="N">  <responseType>Closed with monetary relief</responseType>  </response>  </complaint>  <complaint id="2364257">  <event type="received" date="2017-02-28"/>  <event type="sentToCompany" date="2017-02-28"/>  <product>  <productType>Credit card</productType>  </product>  <issue>  <issueType>Other fee</issueType>  </issue>  <consumerNarrative>Was a happy XXXX card member for years, in late XX/XX/2016 XXXX converted  the card portfolio to Barclaycard ( XXXX ). We almost never carry a balance over, but we  started to in XX/XX/XXXX and Barclay has been overcharging the interest expense every  month. Instead of charging interest on the carried balance they charged it on the entire  average balance. So if we charged {$3000.00} last month and carried {$3000.00} from  previous months then they charged us 15 % of the {$6000.00} = {$75.00}, should have been  {$37.00} in interest charges. They are double dipping, getting the interchange fee ( 1.5  % of purchase, equal to an 18 % apr ), plus they are getting the interest on the  purchases at 15 %, that is the equivalent of an 33 % interest charge. I feel this  practice is very unethical, if not illegal. We converted, not by our choice, from XXXX  to Barclaycard MasterCard, so if we leave we lose all the points we acquired in previous  years. Completely unfair and is why the big financials have the hated reputation they  have now. Hope you folks over there can investigate.</consumerNarrative>  <company>  <companyName>BARCLAYS BANK DELAWARE</companyName>  <companyState>MA</companyState>  <companyZip>19904</companyZip>  </company>  <submitted via="Web"/>  <response timely="Y" consumerDisputed="Y">  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to  provide a public response</publicResponse>  <responseType>Closed with explanation</responseType>  </response>  </complaint>  <complaint id="2327502">  <event type="received" date="2017-02-03"/>  <event type="sentToCompany" date="2017-02-03"/>  <product>  <productType>Credit reporting</productType>  </product>  <issue>  <issueType>Incorrect information on credit report</issueType>  <subissue>Account status</subissue>  </issue>  <consumerNarrative>Checked my credit report after filing complaint with CFPB on XXXX. Was  finally able to get access to the dispute forms and the XXXX XXXX account scheduled for  deletion XX/XX/XXXX2017 was still on record. After already registering with my report  number, name and social security and placing the dispute in the " dispute cart '', when  I attempted to upload as instructed, I was taken to another form which requested the  same ( and more ) information which was already a matter of record in order to get  access to the report in the first place. Screenshots attached. Designed to  discourage?</consumerNarrative>  <company>  <companyName>Experian Information Solutions Inc.</companyName>  <companyState>NY</companyState>  <companyZip>10020</companyZip>  </company>  <submitted via="Web"/>  <response timely="Y" consumerDisputed="N">  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to  provide a public response</publicResponse>  <responseType>Closed with non-monetary relief</responseType>  </response>  </complaint>  <complaint id="2356421">  <event type="received" date="2018-02-23"/>  <event type="sentToCompany" date="2018-02-23"/>  <product>  <productType>Bank account or service</productType>  <subproduct>Savings account</subproduct>  </product>  <issue>  <issueType>Deposits and withdrawals</issueType>  </issue>  <consumerNarrative>I deposited what turned out to be a fraudulent check drawn on Wells Fargo  by mobile deposit to my savings account at Wells Fargo on XXXX at XXXX XXXX Time for  {$2400.00}. They gave me full availability of the {$2400.00} on XXXX at which time I  withdrew {$2200.00} and the bank then returned the deposited check on XXXX creating an  overdraft in my account of over {$2000.00}. Wells Fargo rep explained that they do not  process mobile deposits until late the night one day after the deposit was made. This  means they honored the withdrawal request before they processed the transaction. That  gave me the false assurance that the deposited check was good. The cash is gone to the  perpetrator and now they want me to cover the overdraft. The fact they wait a whole  business day before processing these deposits is for their convenience and the consumer  should not be held accountable for the consequences of this delay. Also UCC 4-301 ( b )  addresses the final payment of on-us checks deposited and states that the payor bank has  until midnight of the next banking day to decide whether to honor the check. If they do  n't act by midnight deadline, they lose the right to dishonor the check. 4-214 ( c ),  4-301 ( b ). The mobile deposit confirmation states " The following mobile deposit was  made on XXXX at XXXX Time '' and her account statement shows the deposit under the  posting date of XXXX. Therefore, applying UCC 4-214 ( c ), the deposited check drawn on  Wells Fargo should have been returned and charged back under the posting date of XXXX.  It was not. The chargeback is posted under processing date of XXXX.</consumerNarrative>  <company>  <companyName>Wells Fargo &amp; Company</companyName>  <companyState>AZ</companyState>  <companyZip>85043</companyZip>  </company>  <submitted via="Web"/>  <response timely="N" consumerDisputed="N">  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to  provide a public response</publicResponse>  <responseType>Closed with explanation</responseType>  </response>  </complaint>  <complaint id="2112558">  <event type="received" date="2016-09-15"/>  <event type="sentToCompany" date="2016-09-15"/>  <product>  <productType>Debt collection</productType>  <subproduct>Medical</subproduct>  </product>  <issue>  <issueType>Continued attempts to collect debt not owed</issueType>  <subissue>Debt is not mine</subissue>  </issue>  <consumerNarrative>I am a veteran widow whom is a recipient of Maryland State Medicaid and  have been for several years. Therefore, the State is responsible for my health bills at  XXXX cost to me.</consumerNarrative>  <company>  <companyName>Round Two Recovery</companyName>  <companyState>OK</companyState>  <companyZip>73135</companyZip>  </company>  <submitted via="Web"/>  <response timely="N" consumerDisputed="N">  <responseType>Untimely response</responseType>  </response>  </complaint>  <complaint id="837784">  <event type="received" date="2014-05-05"/>  <event type="sentToCompany" date="2014-05-06"/>  <product>  <productType>Student loan</productType>  <subproduct>non-federal student loan</subproduct>  </product>  <issue>  <issueType>Dealing with my lender or service</issueType>  <subissue>Need information about my balance/terms</subissue>  </issue>  <company>  <companyName>Navient Solutions, LLC</companyName>  <companyState>DE</companyState>  <companyZip>19802</companyZip>  </company>  <submitted via="Web"/>  <response timely="Y" consumerDisputed="N">  <responseType>Closed with monetary relief</responseType>  </response>  </complaint>  <complaint id="14038">  <company>  <companyName>U.S. BANCORP</companyName>  <companyState>AZ</companyState>  <companyZip>85008</companyZip>  </company>  <event type="sentToCompany" date="2012-01-22"/>  <submitted via="Referral"/>  <issue>  <issueType>Loan servicing, payments, escrow account</issueType>  </issue>  <product>  <productType>Mortgage</productType>  <subproduct>Other mortgage</subproduct>  </product>  <event type="received" date="2012-01-17"/>  <response timely="Y" consumerDisputed="Y">  <responseType>Closed without relief</responseType>  </response>  </complaint>  </consumerComplaints> |

Figure 1.2.1 File A with embedded DTD

### 1.2.1 .1 XML Prose for element, attribute, and attribute values for file A

The root element is consumerComplaints. The consumerComplaints element is the parent node for complaint elements.

The complaint provides information pertaining to a complaint received by the legacy system to track from customers complaints. The element complaint contains children elements including either:

* The occurrence of 2 event elements, a product element, an issue element, zero or 1 occurrences of a consumerNarrative, a company, zero or 1 occurrences of submitted, and a response element.
* A company element, an event element, a submitted element, issue element, product element, an event element, and finally a response element.

|  |
| --- |
| <!ELEMENT complaint ((event, event, product, issue, consumerNarrative?, company, submitted?,response) | (company, event, submitted, issue, product, event, response)) |

The complaint element has the fixed required attribute of an id. It was considered that the id could be of type ID, however, that is reserved for unique names which cannot start with an integer value, so a suitable alternative was to use an NMTOKEN which would at minimum ensure no whitespaces are allowed in the id value, however this does not ensure uniqueness of the ID. Placing a character in front of the value (example: id=12345 could be id=M12345) would have ensured uniqueness, this task within the project was creating a DTD for the file as is, so NMTOKEN was an appropriate attribute type.

In order to validate the xml file with a DTD it should be noted, that order does matter, which is why there are two different xml element options to define a complaint element.

event is an empty element with attributes that provide information regarding dates and context to the dates – indicating if a date is in regard to when the information was “received”, or when the information was “sentToCompany”. The context for the date is stored in a required attribute called type. The date is stored in a required attribute named date. Regardless of order, the event element will be found exactly twice in a complaint element. It will either be the first two items, or the second, and the sixth element found within a complaint.

product is an element providing product information for a consumer complaint. The product element occurs exactly once in the complaint element. The product element will contain exactly one productType and will include 0 or 1 occurrences of a subproduct.

issue is an element providing issue information for a given complaint. An issue element will contain exactly one child element of issueType and will include 0-1 occurrences of a subissue.

consumerNarrative is an element appears to be a narrative of details regarding the complaint. This element will occur 0-1 times in the complaint element provided the element event is the first element found within a complaint. If event is not the first element in a complaint, based on the sample data provided, it will not be included in the complaint element.

company is an element that holds data pertaining to the company that the complaint is filed against. company occurs exactly once within a complaint. The company element includes elements companyName, companyState, and companyZip which each occur exactly once within the company element.

submitted is an element indicating how a complaint was provided. submitted is an empty element which includes an attribute via which can hold the possible values of Phone, Referral or Web. submitted is an element that will occur 1 time in a complaint element.

response is an element that provides information regarding how a response was handled. response occurs exactly once within a complaint element. The response element includes a publicResponse element that will occur 0-1 times within a response as well as a responseType which will occur exactly 1 time within a complaint element. The response element includes a required attribute timely which can be either a Y or N. The response element also includes a required attribute consumerDisputed which can either have a value of Y or N.

No default values were provided to attributes since all attributes were determined to be required.

### 1.2.1 .2 Difficult decisions in designing DTD for file A XML

Standard DTD convention would be to claim complaint element should be defined as follows below:

|  |
| --- |
| <!ELEMENT complaint ((event+, product, issue, consumerNarrative?, company, submitted,response) | (company, event, submitted, issue, product, event, response)) |

Figure 1.2.1.2.1

However, it was noted that there are always 2 events, so in the DTD the choice was made to explicitly call out that there are two events rather than a minimum of 1.

The last XML element, the complaint with an id of 14038 posed interesting choices to be made.

For id 14038 the element follows the convention of:

|  |
| --- |
| (company, event, submitted, issue, product, event, response) |

Figure 1.2.1.2.2

Given order does matter in the DTD to pass validation, it is concerning that this element does not contain data for a consumerNarrative so we have no indication of where it may or may not be in an element that does follow this ordering format (referred to as ordering format #2) in other sample data. In addition, since we only have this one data element, it does appear that the submitted is required, when that may be due to the fact there is only one data element that follows this ordering format #2. It was decided to not make the submitted optional, even though it may very well be. It was also determined not to include consumerNarrative since it would be unclear where it would fit in this new ordering. The decision was made to follow the data, and this design decision would require follow-up for either more sample data, or gather requirements from the support team of system A to ensure this DTD could be applied to all data coming from System A. It is very likely that it will require alterations.

### 1.2.2 File B XML Document DTD Implementation

Below is the embedded DTD implementation for file B. The file has been validated using the website: <http://xmlvalidator.new-studio.org/> with a screen shot of its validity found in [Figure Appendix.1.2.2 File B Embedded DTD Validation](#_Figure_Appendix.1.2.2_File)

The file B with an embedded DTD is included in this solution folder structure under:

[\File\_B\_with\_Embedded DTD\File\_B\_with\_Embedded.xml](File_B_with_Embedded%20DTD/File_B_with_Embedded.xml)

The file B with an external DTD is included in this solution folder structure under:

[\File\_B\_with\_External\_DTD\Consumer\_Complaints\_FileB\_with\_External\_DTD.xml](File_B_with_External_DTD/Consumer_Complaints_FileB_with_External_DTD.xml)

The external DTD was validated using XMLSpy.

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8"?>  <!DOCTYPE consumerComplaints [  <!ENTITY redaction "XXXX">  <!ELEMENT consumerComplaints (complaint+)>  <!ELEMENT complaint (  (event, event, product, issue, consumerNarrative?, company, submitted?, response)| (company, event, issue, product, event, response))>  <!ELEMENT event EMPTY>  <!ELEMENT product (productType, subproduct?)>  <!ELEMENT productType (#PCDATA)>  <!ELEMENT subproduct (#PCDATA)>  <!ELEMENT issue (issueType, subissue?)>  <!ELEMENT issueType (#PCDATA)>  <!ELEMENT subissue (#PCDATA)>  <!ELEMENT company (companyName,companyState,companyZip)>  <!ELEMENT companyName (#PCDATA)>  <!ELEMENT companyState (#PCDATA)>  <!ELEMENT companyZip (#PCDATA)>  <!ELEMENT submitted EMPTY>  <!ELEMENT response (publicResponse?, responseType)>  <!ELEMENT responseType (#PCDATA)>  <!ELEMENT publicResponse (#PCDATA)>  <!ELEMENT consumerNarrative (#PCDATA)>  <!ATTLIST complaint id NMTOKEN #REQUIRED>  <!ATTLIST complaint submissionType (Phone|Referral|Web) #IMPLIED>  <!ATTLIST event type (received | sentToCompany) #REQUIRED >  <!ATTLIST event date CDATA #REQUIRED>  <!ATTLIST response timely (yes|no) #IMPLIED>  <!ATTLIST response consumerDisputed (Y|N) #REQUIRED>  ]>  <consumerComplaints>  <complaint id="759222" submissionType="Referral">  <event type="received" date="2014-03-12"/>  <event type="sentToCompany" date="2014-03-17"/>  <product>  <productType>Mortgage</productType>  <subproduct>Other mortgage</subproduct>  </product>  <issue>  <issueType>Loan modification,collection,foreclosure</issueType>  </issue>  <company>  <companyName>M&amp;T Bank Corporation</companyName>  <companyState>MI</companyState>  <companyZip>48382</companyZip>  </company>  <response timely="yes" consumerDisputed="Y">  <responseType>Closed with explanation</responseType>  </response>  </complaint>  <complaint id="596562" submissionType="Phone">  <event date="2013-11-13" type="received"/>  <event type="sentToCompany" date="2013-11-20"/>  <product>  <productType>Mortgage</productType>  <subproduct>Conventional adjustable mortgage</subproduct>  </product>  <issue>  <issueType> Loan servicing, payments, escrow account</issueType>  </issue>  <company>  <companyName>U.S. BANCORP </companyName>  <companyState>MN</companyState>  <companyZip>48322</companyZip>  </company>  <response consumerDisputed="N" timely="yes">  <responseType>Closed with monetary relief</responseType>  </response>  </complaint>  <complaint id="2364257">  <event date="2017-02-28" type="received"/>  <event type="sentToCompany" date="2017-02-28"/>  <product>  <productType>Credit card</productType>  </product>  <issue>  <issueType>Other fee</issueType>  </issue>  <consumerNarrative>Was a happy &redaction; card member for years, in late XX/XX/2016 &redaction; converted  the card portfolio to Barclaycard ( &redaction; ). We almost never carry a balance over, but we  started to in XX/XX/XXXX and Barclay has been overcharging the interest expense every  month. Instead of charging interest on the carried balance they charged it on the entire  average balance. So if we charged {$3000.00} last month and carried {$3000.00} from  previous months then they charged us 15 % of the {$6000.00} = {$75.00}, should have been  {$37.00} in interest charges. They are double dipping, getting the interchange fee ( 1.5  % of purchase, equal to an 18 % apr ), plus they are getting the interest on the  purchases at 15 %, that is the equivalent of an 33 % interest charge. I feel this  practice is very unethical, if not illegal. We converted, not by our choice, from &redaction;  to Barclaycard MasterCard, so if we leave we lose all the points we acquired in previous  years. Completely unfair and is why the big financials have the hated reputation they  have now. Hope you folks over there can investigate.</consumerNarrative>  <company>  <companyName>BARCLAYS BANK DELAWARE</companyName>  <companyState>MA</companyState>  <companyZip>19904</companyZip>  </company>  <submitted/>  <response consumerDisputed="Y" timely="yes">  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to provide a public response</publicResponse>  <responseType>Closed with explanation</responseType>  </response>  </complaint>  <complaint id="2327502" submissionType="Web">  <event type="received" date="2017-02-03"/>  <event type="sentToCompany " date="2017-02-03"/>  <product>  <productType>Credit reporting</productType>  </product>  <issue>  <issueType>Incorrect information on credit report</issueType>  <subissue>Account status</subissue>  </issue>  <consumerNarrative>Checked my credit report after filing complaint with CFPB on &redaction;. Was  finally able to get access to the dispute forms and the &redaction; &redaction; account scheduled for  deletion XX/XX/XXXX2017 was still on record. After already registering with my report  number, name and social security and placing the dispute in the " dispute cart '', when  I attempted to upload as instructed, I was taken to another form which requested the  same ( and more ) information which was already a matter of record in order to get  access to the report in the first place. Screenshots attached. Designed to  discourage?</consumerNarrative>  <company>  <companyName>Experian Information Solutions Inc.</companyName>  <companyState>NY</companyState>  <companyZip>10020</companyZip>  </company>  <response timely="yes" consumerDisputed="N">  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to  provide a public response</publicResponse>  <responseType>Closed with non-monetary relief</responseType>  </response>  </complaint>  <complaint id="2356421" submissionType="Web">  <event type="received" date="2018-02-23"/>  <event type="sentToCompany " date="2018-02-23"/>  <product>  <productType>Bank account or service</productType>  <subproduct>Savings account</subproduct>  </product>  <issue>  <issueType>Deposits and withdrawals</issueType>  </issue>  <consumerNarrative>I deposited what turned out to be a fraudulent check drawn on Wells Fargo  by mobile deposit to my savings account at Wells Fargo on &redaction; at &redaction; &redaction; Time for  {$2400.00}. They gave me full availability of the {$2400.00} on &redaction; at which time I  withdrew {$2200.00} and the bank then returned the deposited check on &redaction; creating an  overdraft in my account of over {$2000.00}. Wells Fargo rep explained that they do not  process mobile deposits until late the night one day after the deposit was made. This  means they honored the withdrawal request before they processed the transaction. That  gave me the false assurance that the deposited check was good. The cash is gone to the  perpetrator and now they want me to cover the overdraft. The fact they wait a whole  business day before processing these deposits is for their convenience and the consumer  should not be held accountable for the consequences of this delay. Also UCC 4-301 ( b )  addresses the final payment of on-us checks deposited and states that the payor bank has  until midnight of the next banking day to decide whether to honor the check. If they do  n't act by midnight deadline, they lose the right to dishonor the check. 4-214 ( c ),  4-301 ( b ). The mobile deposit confirmation states " The following mobile deposit was  made on &redaction; at &redaction; Time '' and her account statement shows the deposit under the  posting date of &redaction;. Therefore, applying UCC 4-214 ( c ), the deposited check drawn on  Wells Fargo should have been returned and charged back under the posting date of &redaction;.  It was not. The chargeback is posted under processing date of &redaction;.</consumerNarrative>  <company>  <companyName>Wells Fargo &amp; Company</companyName>  <companyState>AZ</companyState>  <companyZip>85043</companyZip>  </company>  <response timely="no" consumerDisputed="N">  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to  provide a public response</publicResponse>  <responseType>Closed with explanation</responseType>  </response>  </complaint>  <complaint id="2112558" submissionType="Web">  <event type="received" date="2016-09-15"/>  <event type="sentToCompany " date="2016-09-15"/>  <product>  <productType>Debt collection</productType>  <subproduct>Medical</subproduct>  </product>  <issue>  <issueType>Continued attempts to collect debt not owed</issueType>  <subissue>Debt is not mine</subissue>  </issue>  <consumerNarrative>I am a veteran widow whom is a recipient of Maryland State Medicaid and  have been for several years. Therefore, the State is responsible for my health bills at  &redaction; cost to me.</consumerNarrative>  <company>  <companyName>Round Two Recovery</companyName>  <companyState>OK</companyState>  <companyZip>73135</companyZip>  </company>  <response timely="no" consumerDisputed="N">  <responseType>Untimely response</responseType>  </response>  </complaint>  <complaint id="837784">  <!-- Note: Sally modified this event on 2014-05-06 -->  <event type="received" date="2014-05-05"/>  <event date="2014-05-06" type="sentToCompany"/>  <product>  <productType>Student loan</productType>  <subproduct>non-federal student loan</subproduct>  </product>  <issue>  <issueType>Dealing with my lender or service</issueType>  <subissue>Need information about my balance/terms</subissue>  </issue>  <company>  <companyName>Navient Solutions, LLC</companyName>  <companyState>DE</companyState>  <companyZip>19802</companyZip>  </company>  <response consumerDisputed="N">  <responseType>Closed with monetary relief</responseType>  </response>  </complaint>  <complaint id="14038" submissionType="Referral">  <company>  <companyName>U.S. BANCORP</companyName>  <companyState>AZ</companyState>  <companyZip>85008</companyZip>  </company>  <event type="sentToCompany" date="2012-01-22"/>  <issue>  <issueType>Loan servicing, payments, escrow account</issueType>  </issue>  <product>  <productType>Mortgage</productType>  <subproduct>Other mortgage</subproduct>  </product>  <event type="received" date="2012-01-17"/>  <response consumerDisputed="Y">  <responseType>Closed without relief</responseType>  </response>  </complaint>  </consumerComplaints> |

### 1.2.2.1 XML Prose for element, attribute, and attribute values for file B

The root element is consumerComplaints. The consumerComplaints element is the parent node for complaint elements.

The complaint provides information pertaining to a complaint received by the system to process. The element complaint contains children elements including either:

* Two occurrences of an event, a product element, and issue element, 0-1 occurrences of a consumerNarrative , a company element, 0-1 occurrences of a submitted element , and a response element
* A company element, an event element, an issue element, a product element, and event element and a response element.

|  |
| --- |
| <!ELEMENT complaint (  (event, event, product, issue, consumerNarrative?, company, submitted?, response)| (company, event, issue, product, event, response))> |

The complaint element has the fixed required attribute of an id. It was considered that the id could be of type ID, however, that is reserved for unique names which cannot start with an integer value, so a suitable alternative was to use an NMTOKEN which would at minimum ensure no whitespaces could be allowed in the id value, however this does not ensure uniqueness of the ID. Placing a character in front of the value (example: id=12345 could be id=M12345) would have ensured uniqueness, but we are creating a DTD for the file as is, so NMTOKEN was an appropriate attribute type.

The complaint element has an implied attribute of a submissionType which has possible values of Phone, Referral, or Web. The attribute is not in all complaint elements which is why it is an implied value.

To validate the xml file with a DTD it should be noted, that order does matter, which is why there are two different xml element options to define a complaint element.

event is an empty element with attributes that provide information regarding dates and context to the dates – indicating if a date is in regard to when the information was “received”, or when the information was “sentToCompany” and “sentToCompany “, note the extra space at the end of the value. Even though this was found, it does pass validation of the DTD. The context for the date is stored in a required attribute called type. The date is stored in a required attribute named date. Regardless of order, the event element will be found exactly twice in a complaint element. It will either be the first two items, or the second, and the fifth element found within a complaint.

product is an element providing product information for a consumer complaint. The product element occurs exactly once in the complaint element. The product element will contain exactly once within the complaint element. A product element will include a productType and will include 0 or 1 occurrences of a subproduct. Product will either be the 3rd or the 4th child element found within a complaint.

issue is an element providing issue information for a given complaint. An issue element will contain exactly one child element of issueType and will include 0-1 occurrences of a subissue. issue will either be the 3rd of 4th child element found within a complaint.

consumerNarrative is an element appears to be a narrative of details regarding the complaint. This element will occur 0-1 times in the complaint element provided a complaint begins with 2 event elements.

company is an element that holds data pertaining to the company that the complaint is filed against. company occurs exactly once within a complaint. The company element includes elements companyName, companyState, and companyZip which each occur exactly once within the company element.

submitted is an empty element that will occur 0-1 times within a complaint. It is unclear what information this element is providing provided the complaint begins with an event. In the case that a complaint begins with company, submitted will not be included in the complaint.

response is an element that provides information regarding how a response was handled. response occurs exactly once within a complaint element. The response element includes a publicResponse element that will occur 0-1 times within a response as well as a responseType. A response will occur exactly 1 time within a complaint element. The response element includes an implied attribute timely which can be either a yes or no. The response element also includes a required attribute consumerDisputed which can either have a value of Y or N.

No default values were provided in the DTD due to the fact that it would be difficult to determine with a single file what the implied value should be and would result in changing the output of the XML file.

It could be noted that for id: 837784 the response attribute of timely in File A was determined to be Y, and given this value is not included in File B, it could be considered that the default value for timely should indicate a value of yes which would be required for the two files to be equivalent.

### 1.2.2.2 Difficult decisions in designing DTD for file B XML

|  |
| --- |
| <!ELEMENT complaint ((event+, product, issue, ((company, response) | (consumerNarrative, company, submitted?, response)) ) | (company, event, issue, product, event, response))> |
| <!ELEMENT complaint ((event+, product, issue, consumerNarrative?, ((company, response) | (company, submitted?, response))) | (company, event, issue, product, event, response))> |
| <!ELEMENT complaint ( (event+, product, issue, consumerNarrative?, company, submitted?, response)| (company, event, issue, product, event, response))> |

Table 1.2.2.2

Table 1.2.2.2 displays 3 additional DTD definitions for a complaint that pass validation. The first design decision that was made to modify the event+ to be event, event. This ensures that only 2 events can be defined as the DTD definition of an event+ would allow for more than 2 events, and all the complaint data elements had either 2 events, so this will help with ensuring data integrity.

## Part 1.3 Canonicalize the two data files and run checksums again to check for equivalence

According to Wikipedia (https://en.wikipedia.org/wiki/Canonical\_XML, 2016), Canonical XML is “intended to allow relatively simple comparison of pairs of XML documents for equivalence”, it further states that:, “in a special context users might care about special semantics beyond the generic logical equivalence with which Canonical XML is associated” which is the case for this project, and further it states that, “Such contexts are beyond the scope of Canonical XML.”

Armed with the information that Canonical XML is a great starting point, through analysis of the datasets, it was found that strictly providing a canonical XML file is not be sufficient to determine if the two files can be determined to be identical. Below is the analysis indicating why strictly performing the canonical process on the 2 files will not be sufficient for determining if these 2 files are equivalent.

To evaluate the data found within the two files for logical equivalence, XML canonicalization is not sufficient, as the analysis provided below details.

The information following in this section is a review of why a strict canonicalization of the XML files is not suffieint.

Figure 1.3.1 shows that a canonicalized XML file will retain whitespaces as outlined by W3C..

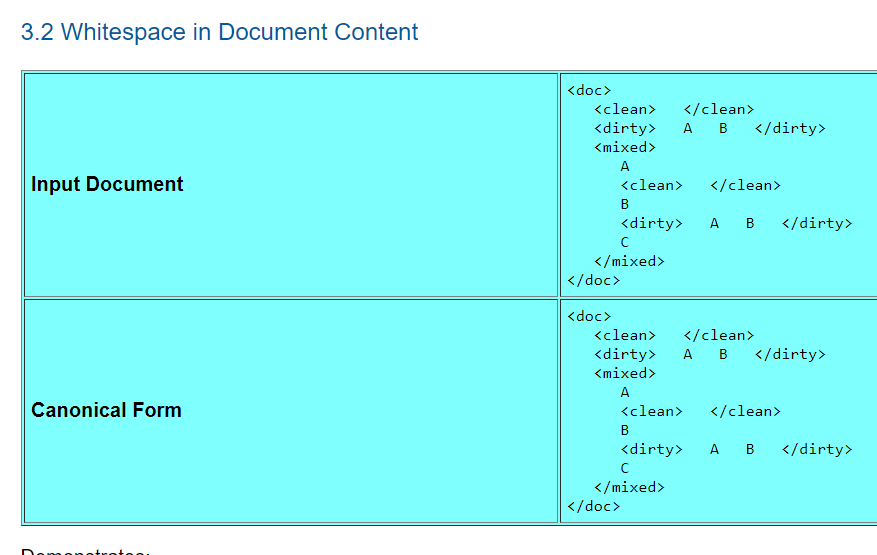


Figure 1.3.1 W3C Figure (W3C, n.d.)

**Violation**: According to W3C, it is required to retain all whitespace between consecutive start tags. For complaint ID: **596562**, if we strictly followed the XML canonicalization process, the files would be evaluated as not equivalent. Figure 1.3.2 shows file A has no leading spaces, and as Figure 1.3.3 shows file B has leading spaces.

|  |
| --- |
|  |

Figure 1.3.2 File A issueType whitespace between tags

|  |
| --- |
|  |

Figure 1.3.3 File B issueType whitespace between tags

**Potential Violation:** According to W3C, when generating a canionical XML file, a determination can be made if comments are to be saved, or not saved. It should be noted that a comment was found in File B, and not in file A. In order to canonicalize this file, it would be required to not include comments given it is in File B, and not in File A as shown in Figure 1.3.4

|  |
| --- |
|  |

Figure 1.3.4 File B

**Violation**: Retain all whitespace between consecutive start tags. For complaint ID: **596562**, if we strictly followed the XML canonicalization process, the files would be evaluated as not equivalent. It should also be noted that this could result in poor querying of results given the data is not normalized, if a system was returning results based on the company name, this result may not be returned based on the end users query.

|  |
| --- |
|  |

Figure 1.3.5 File A

|  |
| --- |
|  |

Figure 1.3.6 File B showing extra whitespace.

**Interesting point for analysis**: For complaint ID: **2364257**, in File A, there is no submission type provided, which in File B maps to the attribute via, which in file A is provided as web. This would indicate that the default value for the DTD for file A should include a default value of web.

|  |
| --- |
|  |

Figure 1.3.7 File A for 236457 complaint id

|  |
| --- |
|  |

Figure 1.3.8 File B for 236457 complaint id

Recall that for file B, the submissionType is implied and therefore not required (shown in Figure 1.3.9). If the DTD is updated to have default value of Web, the information found within the files can be considered equivalent, however the schema of the files makes it difficult to evaluate. According the W3C, it can be expected that an XPATH implementation with a processor can be used to assist in the canonicalization process. Instead of and generating an XSLT, a .NET script was generated to perform this task.

|  |
| --- |
| <!ATTLIST complaint submissionType (Phone|Referral|Web) #IMPLIED> |

Figure 1.3.9 File B DTD

**Violation:** For publicResponse the character return is included in File A, but is not included in B.

|  |
| --- |
|  |

Figure 1.3.10 File A publicReponse for 2364257

|  |
| --- |
|  |

Figure 1.3.11 File B publicResponse does not contain character return

## Part 1.4 DTD of Final XML File

Below is the embedded DTD implementation for newly created final file (The output was identical for the transformation process). The file has been validated using the website: <http://xmlvalidator.new-studio.org/> with a screen shot of its validity found in [Figure Appendix.1.4.1 File A Embedded DTD Validation](#_Figure_Appendix_1.4.1)

While both files are identical, for completeness, they can be found with the project structure here:

[\Final\_FileAandFileB\_with\_Embedded\_DTD\FinalFileAWithEmbeddedDTD.xml](Final_FileAandFileB_with_Embedded_DTD/FinalFileAWithEmbeddedDTD.xml)

[\Final\_FileAandFileB\_with\_Embedded\_DTD\FinalFileBWithEmbeddedDTD.xml](Final_FileAandFileB_with_Embedded_DTD/FinalFileBWithEmbeddedDTD.xml)

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8"?>  <!DOCTYPE ConsumerComplaints [  <!ELEMENT ConsumerComplaints (complaint)>  <!ELEMENT complaint (Complaint+)>  <!ELEMENT Complaint (id, product, issue, company, response, consumerNarrative?, submissionType, sentToCompanyDate, receivedDate)>  <!ELEMENT id (#PCDATA)>  <!ELEMENT product (productType, subProduct?)>  <!ELEMENT subProduct (#PCDATA)>  <!ELEMENT productType (#PCDATA)>  <!ELEMENT issue (issueType, subIssue?)>  <!ELEMENT issueType (#PCDATA)>  <!ELEMENT subIssue (#PCDATA)>  <!ELEMENT company (companyName, companyState, companyZip)>  <!ELEMENT companyName (#PCDATA)>  <!ELEMENT companyState (#PCDATA)>  <!ELEMENT companyZip (#PCDATA)>  <!ELEMENT response (timely, consumerDisputed, responseType, publicResponse?)>  <!ELEMENT timely (#PCDATA)>  <!ELEMENT consumerDisputed (#PCDATA)>  <!ELEMENT responseType (#PCDATA)>  <!ELEMENT publicResponse (#PCDATA)>  <!ELEMENT consumerNarrative (#PCDATA)>  <!ELEMENT submissionType (#PCDATA)>  <!ELEMENT sentToCompanyDate (#PCDATA)>  <!ELEMENT receivedDate (#PCDATA)>  ]>  <ConsumerComplaints>  <complaint>  <Complaint>  <id>14038</id>  <product>  <productType>Mortgage</productType>  <subProduct>Other mortgage</subProduct>  </product>  <issue>  <issueType>Loan servicing, payments, escrow account</issueType>  </issue>  <company>  <companyName>U.S. BANCORP</companyName>  <companyState>AZ</companyState>  <companyZip>85008</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>Y</consumerDisputed>  <responseType>Closed without relief</responseType>  </response>  <submissionType>Referral</submissionType>  <sentToCompanyDate>2012-01-22</sentToCompanyDate>  <receivedDate>2012-01-17</receivedDate>  </Complaint>  <Complaint>  <id>2112558</id>  <product>  <productType>Debt collection</productType>  <subProduct>Medical</subProduct>  </product>  <issue>  <issueType>Continued attempts to collect debt not owed</issueType>  <subIssue>Debt is not mine</subIssue>  </issue>  <company>  <companyName>Round Two Recovery</companyName>  <companyState>OK</companyState>  <companyZip>73135</companyZip>  </company>  <response>  <timely>N</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Untimely response</responseType>  </response>  <consumerNarrative>I am a veteran widow whom is a recipient of Maryland State Medicaid and&#xD;  have been for several years. Therefore, the State is responsible for my health bills at&#xD;  XXXX cost to me.</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2016-09-15</sentToCompanyDate>  <receivedDate>2016-09-15</receivedDate>  </Complaint>  <Complaint>  <id>2327502</id>  <product>  <productType>Credit reporting</productType>  </product>  <issue>  <issueType>Incorrect information on credit report</issueType>  <subIssue>Account status</subIssue>  </issue>  <company>  <companyName>Experian Information Solutions Inc.</companyName>  <companyState>NY</companyState>  <companyZip>10020</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with non-monetary relief</responseType>  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to provide a public response</publicResponse>  </response>  <consumerNarrative>Checked my credit report after filing complaint with CFPB on XXXX. Was&#xD;  finally able to get access to the dispute forms and the XXXX XXXX account scheduled for&#xD;  deletion XX/XX/XXXX2017 was still on record. After already registering with my report&#xD;  number, name and social security and placing the dispute in the " dispute cart '', when&#xD;  I attempted to upload as instructed, I was taken to another form which requested the&#xD;  same ( and more ) information which was already a matter of record in order to get&#xD;  access to the report in the first place. Screenshots attached. Designed to&#xD;  discourage?</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2017-02-03</sentToCompanyDate>  <receivedDate>2017-02-03</receivedDate>  </Complaint>  <Complaint>  <id>2356421</id>  <product>  <productType>Bank account or service</productType>  <subProduct>Savings account</subProduct>  </product>  <issue>  <issueType>Deposits and withdrawals</issueType>  </issue>  <company>  <companyName>Wells Fargo &amp; Company</companyName>  <companyState>AZ</companyState>  <companyZip>85043</companyZip>  </company>  <response>  <timely>N</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with explanation</responseType>  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to provide a public response</publicResponse>  </response>  <consumerNarrative>I deposited what turned out to be a fraudulent check drawn on Wells Fargo&#xD;  by mobile deposit to my savings account at Wells Fargo on XXXX at XXXX XXXX Time for&#xD;  {$2400.00}. They gave me full availability of the {$2400.00} on XXXX at which time I&#xD;  withdrew {$2200.00} and the bank then returned the deposited check on XXXX creating an&#xD;  overdraft in my account of over {$2000.00}. Wells Fargo rep explained that they do not&#xD;  process mobile deposits until late the night one day after the deposit was made. This&#xD;  means they honored the withdrawal request before they processed the transaction. That&#xD;  gave me the false assurance that the deposited check was good. The cash is gone to the&#xD;  perpetrator and now they want me to cover the overdraft. The fact they wait a whole&#xD;  business day before processing these deposits is for their convenience and the consumer&#xD;  should not be held accountable for the consequences of this delay. Also UCC 4-301 ( b )&#xD;  addresses the final payment of on-us checks deposited and states that the payor bank has&#xD;  until midnight of the next banking day to decide whether to honor the check. If they do&#xD;  n't act by midnight deadline, they lose the right to dishonor the check. 4-214 ( c ),&#xD;  4-301 ( b ). The mobile deposit confirmation states " The following mobile deposit was&#xD;  made on XXXX at XXXX Time '' and her account statement shows the deposit under the&#xD;  posting date of XXXX. Therefore, applying UCC 4-214 ( c ), the deposited check drawn on&#xD;  Wells Fargo should have been returned and charged back under the posting date of XXXX.&#xD;  It was not. The chargeback is posted under processing date of XXXX.</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2018-02-23</sentToCompanyDate>  <receivedDate>2018-02-23</receivedDate>  </Complaint>  <Complaint>  <id>2364257</id>  <product>  <productType>Credit card</productType>  </product>  <issue>  <issueType>Other fee</issueType>  </issue>  <company>  <companyName>BARCLAYS BANK DELAWARE</companyName>  <companyState>MA</companyState>  <companyZip>19904</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>Y</consumerDisputed>  <responseType>Closed with explanation</responseType>  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to provide a public response</publicResponse>  </response>  <consumerNarrative>Was a happy XXXX card member for years, in late XX/XX/2016 XXXX converted&#xD;  the card portfolio to Barclaycard ( XXXX ). We almost never carry a balance over, but we&#xD;  started to in XX/XX/XXXX and Barclay has been overcharging the interest expense every&#xD;  month. Instead of charging interest on the carried balance they charged it on the entire&#xD;  average balance. So if we charged {$3000.00} last month and carried {$3000.00} from&#xD;  previous months then they charged us 15 % of the {$6000.00} = {$75.00}, should have been&#xD;  {$37.00} in interest charges. They are double dipping, getting the interchange fee ( 1.5&#xD;  % of purchase, equal to an 18 % apr ), plus they are getting the interest on the&#xD;  purchases at 15 %, that is the equivalent of an 33 % interest charge. I feel this&#xD;  practice is very unethical, if not illegal. We converted, not by our choice, from XXXX&#xD;  to Barclaycard MasterCard, so if we leave we lose all the points we acquired in previous&#xD;  years. Completely unfair and is why the big financials have the hated reputation they&#xD;  have now. Hope you folks over there can investigate.</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2017-02-28</sentToCompanyDate>  <receivedDate>2017-02-28</receivedDate>  </Complaint>  <Complaint>  <id>596562</id>  <product>  <productType>Mortgage</productType>  <subProduct>Conventional adjustable mortgage</subProduct>  </product>  <issue>  <issueType>Loan servicing, payments, escrow account</issueType>  </issue>  <company>  <companyName>U.S. BANCORP</companyName>  <companyState>MN</companyState>  <companyZip>48322</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with monetary relief</responseType>  </response>  <submissionType>Phone</submissionType>  <sentToCompanyDate>2013-11-20</sentToCompanyDate>  <receivedDate>2013-11-13</receivedDate>  </Complaint>  <Complaint>  <id>759222</id>  <product>  <productType>Mortgage</productType>  <subProduct>Other mortgage</subProduct>  </product>  <issue>  <issueType>Loan modification,collection,foreclosure</issueType>  </issue>  <company>  <companyName>M&amp;T Bank Corporation</companyName>  <companyState>MI</companyState>  <companyZip>48382</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>Y</consumerDisputed>  <responseType>Closed with explanation</responseType>  </response>  <submissionType>Referral</submissionType>  <sentToCompanyDate>2014-03-17</sentToCompanyDate>  <receivedDate>2014-03-12</receivedDate>  </Complaint>  <Complaint>  <id>837784</id>  <product>  <productType>Student loan</productType>  <subProduct>non-federal student loan</subProduct>  </product>  <issue>  <issueType>Dealing with my lender or service</issueType>  <subIssue>Need information about my balance/terms</subIssue>  </issue>  <company>  <companyName>Navient Solutions, LLC</companyName>  <companyState>DE</companyState>  <companyZip>19802</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with monetary relief</responseType>  </response>  <submissionType>Web</submissionType>  <sentToCompanyDate>2014-05-06</sentToCompanyDate>  <receivedDate>2014-05-05</receivedDate>  </Complaint>  </complaint>  </ConsumerComplaints> |

Figure 1.4.1 Final File A with embedded DTD

In addition strictly for completeness given the two files are identical, the figure below is the Final File B with embedded DTD

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8"?>  <!DOCTYPE ConsumerComplaints [  <!ELEMENT ConsumerComplaints (complaint)>  <!ELEMENT complaint (Complaint+)>  <!ELEMENT Complaint (id, product, issue, company, response, consumerNarrative?, submissionType, sentToCompanyDate, receivedDate)>  <!ELEMENT id (#PCDATA)>  <!ELEMENT product (productType, subProduct?)>  <!ELEMENT subProduct (#PCDATA)>  <!ELEMENT productType (#PCDATA)>  <!ELEMENT issue (issueType, subIssue?)>  <!ELEMENT issueType (#PCDATA)>  <!ELEMENT subIssue (#PCDATA)>  <!ELEMENT company (companyName, companyState, companyZip)>  <!ELEMENT companyName (#PCDATA)>  <!ELEMENT companyState (#PCDATA)>  <!ELEMENT companyZip (#PCDATA)>  <!ELEMENT response (timely, consumerDisputed, responseType, publicResponse?)>  <!ELEMENT timely (#PCDATA)>  <!ELEMENT consumerDisputed (#PCDATA)>  <!ELEMENT responseType (#PCDATA)>  <!ELEMENT publicResponse (#PCDATA)>  <!ELEMENT consumerNarrative (#PCDATA)>  <!ELEMENT submissionType (#PCDATA)>  <!ELEMENT sentToCompanyDate (#PCDATA)>  <!ELEMENT receivedDate (#PCDATA)>  ]>  <ConsumerComplaints>  <complaint>  <Complaint>  <id>14038</id>  <product>  <productType>Mortgage</productType>  <subProduct>Other mortgage</subProduct>  </product>  <issue>  <issueType>Loan servicing, payments, escrow account</issueType>  </issue>  <company>  <companyName>U.S. BANCORP</companyName>  <companyState>AZ</companyState>  <companyZip>85008</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>Y</consumerDisputed>  <responseType>Closed without relief</responseType>  </response>  <submissionType>Referral</submissionType>  <sentToCompanyDate>2012-01-22</sentToCompanyDate>  <receivedDate>2012-01-17</receivedDate>  </Complaint>  <Complaint>  <id>2112558</id>  <product>  <productType>Debt collection</productType>  <subProduct>Medical</subProduct>  </product>  <issue>  <issueType>Continued attempts to collect debt not owed</issueType>  <subIssue>Debt is not mine</subIssue>  </issue>  <company>  <companyName>Round Two Recovery</companyName>  <companyState>OK</companyState>  <companyZip>73135</companyZip>  </company>  <response>  <timely>N</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Untimely response</responseType>  </response>  <consumerNarrative>I am a veteran widow whom is a recipient of Maryland State Medicaid and&#xD;  have been for several years. Therefore, the State is responsible for my health bills at&#xD;  XXXX cost to me.</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2016-09-15</sentToCompanyDate>  <receivedDate>2016-09-15</receivedDate>  </Complaint>  <Complaint>  <id>2327502</id>  <product>  <productType>Credit reporting</productType>  </product>  <issue>  <issueType>Incorrect information on credit report</issueType>  <subIssue>Account status</subIssue>  </issue>  <company>  <companyName>Experian Information Solutions Inc.</companyName>  <companyState>NY</companyState>  <companyZip>10020</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with non-monetary relief</responseType>  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to provide a public response</publicResponse>  </response>  <consumerNarrative>Checked my credit report after filing complaint with CFPB on XXXX. Was&#xD;  finally able to get access to the dispute forms and the XXXX XXXX account scheduled for&#xD;  deletion XX/XX/XXXX2017 was still on record. After already registering with my report&#xD;  number, name and social security and placing the dispute in the " dispute cart '', when&#xD;  I attempted to upload as instructed, I was taken to another form which requested the&#xD;  same ( and more ) information which was already a matter of record in order to get&#xD;  access to the report in the first place. Screenshots attached. Designed to&#xD;  discourage?</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2017-02-03</sentToCompanyDate>  <receivedDate>2017-02-03</receivedDate>  </Complaint>  <Complaint>  <id>2356421</id>  <product>  <productType>Bank account or service</productType>  <subProduct>Savings account</subProduct>  </product>  <issue>  <issueType>Deposits and withdrawals</issueType>  </issue>  <company>  <companyName>Wells Fargo &amp; Company</companyName>  <companyState>AZ</companyState>  <companyZip>85043</companyZip>  </company>  <response>  <timely>N</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with explanation</responseType>  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to provide a public response</publicResponse>  </response>  <consumerNarrative>I deposited what turned out to be a fraudulent check drawn on Wells Fargo&#xD;  by mobile deposit to my savings account at Wells Fargo on XXXX at XXXX XXXX Time for&#xD;  {$2400.00}. They gave me full availability of the {$2400.00} on XXXX at which time I&#xD;  withdrew {$2200.00} and the bank then returned the deposited check on XXXX creating an&#xD;  overdraft in my account of over {$2000.00}. Wells Fargo rep explained that they do not&#xD;  process mobile deposits until late the night one day after the deposit was made. This&#xD;  means they honored the withdrawal request before they processed the transaction. That&#xD;  gave me the false assurance that the deposited check was good. The cash is gone to the&#xD;  perpetrator and now they want me to cover the overdraft. The fact they wait a whole&#xD;  business day before processing these deposits is for their convenience and the consumer&#xD;  should not be held accountable for the consequences of this delay. Also UCC 4-301 ( b )&#xD;  addresses the final payment of on-us checks deposited and states that the payor bank has&#xD;  until midnight of the next banking day to decide whether to honor the check. If they do&#xD;  n't act by midnight deadline, they lose the right to dishonor the check. 4-214 ( c ),&#xD;  4-301 ( b ). The mobile deposit confirmation states " The following mobile deposit was&#xD;  made on XXXX at XXXX Time '' and her account statement shows the deposit under the&#xD;  posting date of XXXX. Therefore, applying UCC 4-214 ( c ), the deposited check drawn on&#xD;  Wells Fargo should have been returned and charged back under the posting date of XXXX.&#xD;  It was not. The chargeback is posted under processing date of XXXX.</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2018-02-23</sentToCompanyDate>  <receivedDate>2018-02-23</receivedDate>  </Complaint>  <Complaint>  <id>2364257</id>  <product>  <productType>Credit card</productType>  </product>  <issue>  <issueType>Other fee</issueType>  </issue>  <company>  <companyName>BARCLAYS BANK DELAWARE</companyName>  <companyState>MA</companyState>  <companyZip>19904</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>Y</consumerDisputed>  <responseType>Closed with explanation</responseType>  <publicResponse>Company has responded to the consumer and the CFPB and chooses not to provide a public response</publicResponse>  </response>  <consumerNarrative>Was a happy XXXX card member for years, in late XX/XX/2016 XXXX converted&#xD;  the card portfolio to Barclaycard ( XXXX ). We almost never carry a balance over, but we&#xD;  started to in XX/XX/XXXX and Barclay has been overcharging the interest expense every&#xD;  month. Instead of charging interest on the carried balance they charged it on the entire&#xD;  average balance. So if we charged {$3000.00} last month and carried {$3000.00} from&#xD;  previous months then they charged us 15 % of the {$6000.00} = {$75.00}, should have been&#xD;  {$37.00} in interest charges. They are double dipping, getting the interchange fee ( 1.5&#xD;  % of purchase, equal to an 18 % apr ), plus they are getting the interest on the&#xD;  purchases at 15 %, that is the equivalent of an 33 % interest charge. I feel this&#xD;  practice is very unethical, if not illegal. We converted, not by our choice, from XXXX&#xD;  to Barclaycard MasterCard, so if we leave we lose all the points we acquired in previous&#xD;  years. Completely unfair and is why the big financials have the hated reputation they&#xD;  have now. Hope you folks over there can investigate.</consumerNarrative>  <submissionType>Web</submissionType>  <sentToCompanyDate>2017-02-28</sentToCompanyDate>  <receivedDate>2017-02-28</receivedDate>  </Complaint>  <Complaint>  <id>596562</id>  <product>  <productType>Mortgage</productType>  <subProduct>Conventional adjustable mortgage</subProduct>  </product>  <issue>  <issueType>Loan servicing, payments, escrow account</issueType>  </issue>  <company>  <companyName>U.S. BANCORP</companyName>  <companyState>MN</companyState>  <companyZip>48322</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with monetary relief</responseType>  </response>  <submissionType>Phone</submissionType>  <sentToCompanyDate>2013-11-20</sentToCompanyDate>  <receivedDate>2013-11-13</receivedDate>  </Complaint>  <Complaint>  <id>759222</id>  <product>  <productType>Mortgage</productType>  <subProduct>Other mortgage</subProduct>  </product>  <issue>  <issueType>Loan modification,collection,foreclosure</issueType>  </issue>  <company>  <companyName>M&amp;T Bank Corporation</companyName>  <companyState>MI</companyState>  <companyZip>48382</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>Y</consumerDisputed>  <responseType>Closed with explanation</responseType>  </response>  <submissionType>Referral</submissionType>  <sentToCompanyDate>2014-03-17</sentToCompanyDate>  <receivedDate>2014-03-12</receivedDate>  </Complaint>  <Complaint>  <id>837784</id>  <product>  <productType>Student loan</productType>  <subProduct>non-federal student loan</subProduct>  </product>  <issue>  <issueType>Dealing with my lender or service</issueType>  <subIssue>Need information about my balance/terms</subIssue>  </issue>  <company>  <companyName>Navient Solutions, LLC</companyName>  <companyState>DE</companyState>  <companyZip>19802</companyZip>  </company>  <response>  <timely>Y</timely>  <consumerDisputed>N</consumerDisputed>  <responseType>Closed with monetary relief</responseType>  </response>  <submissionType>Web</submissionType>  <sentToCompanyDate>2014-05-06</sentToCompanyDate>  <receivedDate>2014-05-05</receivedDate>  </Complaint>  </complaint>  </ConsumerComplaints> |

Figure 1.4.2 Final File B with embedded DTD

### 1.4.2.1 XML Prose for element, attribute, and attribute values for final generated file

The root element is consumerComplaints. The consumerComplaints element is the parent node for complaint elements.

The complaint element is the parent node for the Compliant elements. This additional nesting is a direct result of the json implementation.

The Complaint provides information pertaining to a complaint received by either system to process. The element Complaint contains children elements including:

An id, product, issue, company, response, 0-1 occurances of a consumerNarrative, a submissionType, sentToCompanyDate, receivedDate.

An id element is the identifier for a given complaint.

product is an element providing product information for a consumer complaint. The product element occurs exactly once in the complaint element. The product element will contain exactly one productType and will include 0 or 1 occurrences of a subproduct.

issue is an element providing issue information for a given complaint. An issue element will contain exactly one child element of issueType and will include 0-1 occurrences of a subissue.

company is an element that holds data pertaining to the company that the complaint is filed against. company occurs exactly once within a complaint. The company element includes elements companyName, companyState, and companyZip which each occur exactly once within the company element.

response is an element that provides information regarding how a response was handled. response occurs exactly once within a Complaint element. A response includes an element timely, an element comsumerDisputed, an element responseType, and 0-1 occurances of a publicReponse.

consumerNarrative is an element appears to be a narrative of details regarding the complaint. This element will occur 0-1 times in the Complaint.

submissionType is an element that provides information regarding how the submission was collected. subissionType occurs exactly one time in a Complaint element. With the class implementation this value is given a default value of “Web”, and this could be expanded to include an enumeration of the possible types of submission.

sentToCompanyDate is an element that is suspected to be the date the complaint information was sent from the company collecting this information to the consumer of this complaint, which maybe to the company it is against, or perhaps to another organization for analysis. This element will occur exactly one time in a Complaint element.

recievedDate is an element that is suspected to be the date the complaint information was provided to the company collecting this information from a consumer. This element will occur exactly one time in a Complaint element.

### 1.4.1 .2 Difficult decisions in designing DTD for file final XML

In this implementation there really were no difficult decisions that were required to be made. The data was very consistent based on the canonicalization process that was implemented resulting in a consistent data representation which resulted in a simple DTD.

## Appendix

|  |
| --- |
|  |

### Figure Appendix.1.1 Code snipped to compute md5 hash string

|  |
| --- |
|  |

### Figure Appendix.1.2.1 File A Embedded DTD Validation

|  |
| --- |
|  |

### Figure Appendix.1.2.2 File B Embedded DTD Validation

|  |
| --- |
|  |

### Figure Appendix 1.3.1 Winmerge File Comparison

|  |
| --- |
|  |

### Figure Appendix 1.4.1 Embedded DTD Validation

# References

*http://json2csharp.com/*. (n.d.). Retrieved from http://json2csharp.com/: http://json2csharp.com/

*https://en.wikipedia.org/wiki/Canonical\_XML*. (2016, 7 6). Retrieved from https://en.wikipedia.org/wiki/Canonical\_XML: https://en.wikipedia.org/wiki/Canonical\_XML

W3C. (n.d.). *https://www.w3.org/TR/xml-c14n11/*. Retrieved from https://www.w3.org/TR/xml-c14n11/.

*XML Elements vs. Attributes*. (n.d.). Retrieved from w3schools.com: https://www.w3schools.com/xml/xml\_dtd\_el\_vs\_attr.asp

*XmlWriterSettings.NewLineHandling Property*. (n.d.). Retrieved from Microsoft: https://docs.microsoft.com/en-us/dotnet/api/system.xml.xmlwritersettings.newlinehandling?view=netframework-4.7.2