

DTD for File A

```
<!ELEMENT consumerComplaints (complaint+)>
<!ELEMENT complaint (event+, product, Issue, consumerNarrative?, company, submitted,
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 publicResponse (#PCDATA)>
<!ELEMENT responseType (#PCDATA)>

<!ATTLIST complaint id CDATA #REQUIRED>
<!ATTLIST event
    type CDATA #REQUIRED
    date CDATA #REQUIRED>
<!ATTLIST submitted via CDATA #REQUIRED>
<!ATTLIST response
    timely CDATA #REQUIRED
    consumerDisputed CDATA #REQUIRED>
```

DTD for File B

```
<!ELEMENT consumerComplaints (complaint+)>
<!ELEMENT complaint (event+,product,issue,consumerNarrative?,company,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 response (publicResponse?,responseType)>
<!ELEMENT publicResponse (#PCDATA)>
<!ELEMENT responseType (#PCDATA)>

<!ATTLIST complaint
      id      CDATA      #REQUIRED
      submissionType      CDATA      #REQUIRED>
<!ATTLIST event
      type  CDATA      #REQUIRED
      date  CDATA      #REQUIRED>
<!ATTLIST submitted via CDATA #REQUIRED>
<!ATTLIST response
      timely  CDATA      #REQUIRED
      consumerDisputed  CDATA      #REQUIRED>

```

DTD for Final Canonicalized file

```

<!ELEMENT consumerComplaints (complaint+)>
<!ELEMENT complaint (event+, product, Issue, consumerNarrative?, company, submitted,
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 publicResponse (#PCDATA)>
<!ELEMENT responseType (#PCDATA)>

```

```

<!ATTLIST complaint id CDATA #REQUIRED>
<!ATTLIST event
    type CDATA #REQUIRED
    date CDATA #REQUIRED>
<!ATTLIST submitted via CDATA #REQUIRED>
<!ATTLIST response
    timely CDATA #REQUIRED
    consumerDisputed CDATA #REQUIRED>

```

Prose:

File A:

- The root element is consumer complaints which has one or more complaints as child elements under it(hence I have used complaints+).
- Each complaint element contains an id attribute which is a unique identifier. The id attribute is a required value as it could be seen as the primary key in a table.
- Each complaint element further contains child elements which are event, product, Issue, consumerNarrative, company, submitted, response
 - The child element “event” is an empty element with two attributes “type” and “date” which are designed to have a required value. The event element is defined as event+ which means there is an event created for receiving complaint and another for forwarding the complaint to the respective organization.
 - Then there is the product element which contains child element “productType” and “subproduct”. The subproduct is not present in all the complaints, hence is defined as subproduct? In the DTD.
 - The issue element contains two child elements which are “issueType” and “subissue”. A complaint may contain zero or more instances of the “subissue” element.
 - The “consumerNarrative” element is not found in all the complaints. Hence in the DTD it is defined as consumerNarrative?
 - The company element has 3 child elements – “companyName”. “companyState” and “companyZip”. This obviously holds details of the name of the company and its location
 - The “submitted” element which is an empty element but contains a required attributed called “via” which displays the mode of the complaint lodge
 - “Response” element consists of two child elements “publicResponse” and “responseType” and also has two attributes which are “consumerDisputed” and “timely” and they take on binary values

File B:

- The root element is consumer complaints which has one or more complaints as child elements under it(hence I have used complaints+).
- Each complaint element contains an id attribute which is a unique identifier; and a type of submission attribute. The id attribute is a required value as it could be seen as the primary key in a table.
- Each complaint element further contains child elements which are event, product, Issue, consumerNarrative, company, submitted, response
 - The child element “event” is an empty element with two attributes “type” and “date” which are designed to have a required value. The event element is defined as event+ which means there is an event created for receiving complaint and another for forwarding the complaint to the respective organization.
 - Then there is the product element which contains child element “productType” and “subproduct”. The subproduct is not present in all the complaints, hence is defined as subproduct? In the DTD.
 - The issue element contains two child elements which are “issueType” and “subissue”. A complaint may contain zero or more instances of the “subissue” element.
 - The “consumerNarrative” element is not found in all the complaints. Hence in the DTD it is defined as consumerNarrative?
 - The company element has 3 child elements – “companyName”. “companyState” and “companyZip”. This obviously holds details of the name of the company and its location
 - “Response” element consists of two child elements “publicResponse” and “responseType” and also has two attributes which are “consumerDisputed” and “timely” and they take on binary values. Although “timely” is a required attribute, we see it missing for some complaint ids.

Final File:

Since File A is taken as the reference for our canonicalization and integration, the DTD for the final file has the same rules and element/attributes as that of the DTD of file A.