Schemas for Structured Text

- Your data is this *ad-hoc CSV variant*
- Your goal: enforce this format on all such files

```
/foo/bar/data.csv

FIELD1, FIELD2, FIELD3
1, 2, [11,22,33]
4, sym_data, [66, 77]
/a/b/c, 9, 9873AF897FED080989873AF897FED080989873AF897FED0809898
```

Is this CSV variant Well-Formed ?

A DFDL Schema can ensure many things:

- Number of fields in each row matches the number of column headers.
- Column headers are proper identifiers (E.g., all caps alphanumeric)
- Only last column can be variable-length vector or hex blob.
- Fields can be tab or comma separated.
- Fields can have a maximum field length - excluding the vectors/blobs. (which could have a different max length)
- Fields syntax can either match the syntax of integers, identifiers, file names, dates/times, etc., for some list of acceptable field syntaxes.
- Hex blobs are hex-digits only. Enforce maximum length.
- Files obey a specified character-set encoding.
- Maximum number of rows/lines.
- Some characters are disallowed (control characters, for example).

Parse Output Verifies Well-Formed

```
<csv1>
  <version>1.0</version>
  <fileName>/foo/bar/data.csv</fileName>
  <columns>
    <column>FIELD1</COLUMN>
    <column>FIELD2</COLUMN>
    <column>FIELD3</COLUMN>
  </columns>
  <rows>
    <row>
      <c><i>1</i></c><c><i>2</i></c>
      <vector><v>11</v><v>22</v><v>33</v></vector>
    </row>
    <row>
      <c><i>4</i></c><c><s>sym_data</s></c>
      <vector><v>66</v><v>77</v></vector>
    </row>
    <row>
      <c><p>/a/b/c</p></c>
      <c><i>9</i></c>
      <hex>9873AF897FED080989873AF897FED080989873AF897FED0809898</hex>
    </row>
  </rows>
</csv1>
```

Convert to XML - Wrong!

```
<? xml version="1.0" ?>
<textOK><![CDATA[
/foo/bar/data.csv

FIELD1, FIELD2, FIELD3
1, 2, [11,22,33]
4, sym_data, [66, 77]
/a/b/c, 9, 873AF897FED080989873AF897FED080989873AF897FED0809898
]]></textOK>
```

- This is technically valid XML for a trivial schema

  <element name="textOK" type="xs:string"/>

- Schema doesn't do anything. Would be unacceptable or useless for most needs.