

## DCMTK - Bug #1004

### dcm2json produces invalid UTF-8 output for some incorrect DICOM files

2021-08-27 13:00 - Marco Eichelberg

|   |                  |                        |            |
|---|------------------|------------------------|------------|
| <b>Status:</b>  | Closed           | <b>Start date:</b>     | 2021-08-27 |
| <b>Priority:</b>  | Normal           | <b>Due date:</b>       |            |
| <b>Assignee:</b>  | Marco Eichelberg | <b>% Done:</b>         | 100%       |
| <b>Category:</b>  | Library and Apps | <b>Estimated time:</b> | 1:00 hour  |
| <b>Target version:</b>  | 3.6.7            | <b>Compiler:</b>       |            |
| <b>Module:</b>  | dcmdata          |                        |            |
| <b>Operating System:</b>  |                  |                        |            |
| <b>Description</b>  |                  |                        |            |
| <p>The JSON specifications require all JSON scripts to be encoded in UTF-8. dcm2json, therefore, converts DICOM datasets to UTF-8 before writing them to JSON.</p> <p>Currently, however, DICOM files that do not contain (0008,0005) SpecificCharacterSet but do contain extended characters are simply passed through to JSON, possibly resulting in invalid UTF-8.</p> <p>dcm2json should check and report this case, just like dcm2xml or dcmconv +U8.</p> <p>Furthermore, DICOM files not containing (0008,0005) SpecificCharacterSet (or containing the value "ISO_IR 6") should be written to JSON without setting SpecificCharacterSet to ISO_IR 192.</p> <p>Reported 2021-08-26 by Mathieu Malaterre &lt;<a href="mailto:mathieu.malaterre@gmail.com">mathieu.malaterre@gmail.com</a>&gt;.</p> |                  |                        |            |

#### History

##### #1 - 2021-10-01 18:20 - Marco Eichelberg

- Status changed from New to Closed
- % Done changed from 0 to 100
- Estimated time set to 1:00 h

Closed by commit #92da003ff.

#### Files

|            |        |            |                  |
|------------|--------|------------|------------------|
| badUnc.dcm | 513 KB | 2021-08-27 | Marco Eichelberg |
|------------|--------|------------|------------------|