NCO-JSON: JavaScript Object Notation (JSON) for netCDF

Charlie Zender
Departments of Earth System Science and Computer Science, UC Irvine

Acknowledgements:
Chris Barker, Henry Butowsky, Pedro Vicente

ESIP 2018 Summer Meeting
Tucson, AZ
July 17-20, 2018

On Web

A JSON Dialect for netCDF

1. Design
2. Existing ASCII netCDF Formats
3. Existing netCDF-related JSON
4. Namespace
5. Reproducibility
6. Completeness

Design

1. Eight JSON object types: `groups, dimensions, variables, shape, attributes, type, types, data`
2. Access entire netCDF namespace
3. Legible
4. Reproducible
5. Flexible

ASCII netCDF Formats: Common Data Language (CDL)

```
> zender@firn:~$ ncks -v one in.nc
netcdf in {
  variables:
    float one ;
      one:long_name = "one" ;
  data:
    one = 1 ;
} // group /
```

ASCII netCDF Formats: netCDF Markup Language (NcML)

```
> zender@firn:~$ ncks --xml -v one in.nc
<?xml version="1.0" encoding="UTF-8"?>
<ncml:netcdfxmlns:ncml="http://www.unidata.ucar.edu/
namespaces/netcdf/ncml-2.2" location="file:in.nc">
  <ncml:variable name="one" type="float" shape="">
    <ncml:attribute name="long_name" separator="*"
value="one" />
    <ncml:values>1.</ncml:values>
  </ncml:variable>
</ncml:netcdf>
```

Strengths & Weaknesses of ASCII netCDF Formats

	Strengths	Weaknesses
CDL	Legible, Complete	Niche
NcML	XML, THREDDS, OPeNDAP	netCDF4 gaps Verbose

ASCII netCDF Formats: NCO-JSON

```
> zender@firn:~$ ncks --json -v one in.nc
{
  "variables": {
    "one": {
      "type": "float",
      "attributes": {
        "long_name": "one"
      },
      "data": 1.0
    }
  }
}
```

ASCII netCDF Formats: HDF5-JSON

```
> jelenak@thg:~$ h5tojson one.nc
```


```
{
    "apiVersion": "1.1.1",
    "datasets": {
        "f1d21bba-86e3-11e8-83df-760060ca3401": {
            "alias": [
                "/one"
            ],
            "attributes": [
                {
                    "name": "long_name",
                    "shape": {
                        "class": "H5S_SCALAR"
                    },
                    "type": {
                        "charSet": "H5T_CSET_ASCII",
                        "class": "H5T_STRING",
                        "length": 3,
                        "strPad": "H5T_STR_NULLPAD"
                    },
                    "value": "one"
                }
            ],
            "creationProperties": {
                "allocTime": "H5D_ALLOC_TIME_LATE",
                "fillTime": "H5D_FILL_TIME_IFSET",
                "fillValue": 9.969209968386869e+36,
                "layout": {
                    "class": "H5D_CONTIGUOUS"
                }
            },
            "shape": {
                "class": "H5S_SCALAR"
            },
            "type": {
                "base": "H5T_IEEE_F32LE",
                "class": "H5T_FLOAT"
            },
            "value": 1.0
        }
    },
    "groups": {
        "f1d0bac6-86e3-11e8-b54d-760060ca3401": {
            "alias": [
                "/"
            ],
            "attributes": [
                {
                    "name": "_NCProperties",
                    "shape": {
                        "class": "H5S_SCALAR"
                    },
                    "type": {
                        "charSet": "H5T_CSET_ASCII",
                        "class": "H5T_STRING",
                        "length": 57,
                        "strPad": "H5T_STR_NULLPAD"
                    },
                    "value":
"version=1|netcdflibversion=4.4.1.1|hdf5libversion=1.10.2"
                }
            ],
            "links": [
                {
                    "class": "H5L_TYPE_HARD",
                    "collection": "datasets",
                    "id": "f1d21bba-86e3-11e8-83df-760060ca3401",
                    "title": "one"
                }
            ]
        }
    },
    "root": "f1d0bac6-86e3-11e8-b54d-760060ca3401"
}
```

Reproducibility I: Non-pedantic

```
> zender@firn:~$ ncks --json -v one in.nc
{
  "variables": {
    "one": {
      "type": "float",
      "attributes": {
        "long_name": "one"
      },
      "data": 1.0
    }
  }
}
```

Reproducibility II: Pedantic

```
> zender@firn:~$ ncks --jsn_fmt=2 -v one in.nc
{
  "variables": {
    "one": {
      "type": "float",
      "attributes": {
        "long_name": { "type": "char", "data": "one"}
      },
      "data": 1.0
    }
  }
}
```

Reproducibility III: Non-pedantic

```
> zender@firn:~$ ncks --json -v att_var in.nc
...
"attributes": {
   "byte_att": [0, 1, 2, 127, -128, -127, -2, -1],
   "char_att": "Sentence one.\nSentence two.\n",
   "short_att": 37,
   "int_att": 73,
   "float_att": [70.010, 69.0010, 68.010, 67.010],
   "double_att": [70.010, 69.0010, 68.010, 67.0100010],
   "ubyte_att": [0, 1, 2, 127, 128, 254, 255, 0],
   "ushort_att": 37,
   "uint_att": 73,
   "int64_att": 9223372036854775807,
   "uint64_att": 18446744073709551615,
   "string_att": "Hello, World"
},
```

Reproducibility IV: Pedantic

```
> zender@firn:~$ ncks --jsn_fmt=2 -v att_var in.nc
...
"attributes": {
  "byte_att": { "type": "byte", "data": [0, 1, 2, 127, -128, -127, -2,
-1]},
  "char_att": { "type": "char", "data": "Sentence one.\nSentence
two.\n"},
  "short_att": { "type": "short", "data": 37},
  "int_att": { "type": "int", "data": 73},
  "float_att": { "type": "float", "data": [70.010, 69.0010, 68.010,
67.010]},
  "double_att": { "type": "double", "data": [70.010, 69.0010, 68.010,
67.0100010]},
  "ubyte_att": { "type": "ubyte", "data": [0, 1, 2, 127, 128, 254, 255,
0]},
  "ushort_att": { "type": "ushort", "data": 37},
  "uint_att": { "type": "uint64", "data": 73},
  "int64_att": { "type": "int64", "data": 9223372036854775807},
  "uint64_att": { "type": "uint64", "data": 18446744073709551615},
  "string_att": { "type": "string", "data": "Hello, World"}
}
```

Multidimensional Arrays I

```
> ncks -C -H --json -v two_dmn_rec_var in.nc
{
  "dimensions": {
    "lev": 3,
    "time": 10
  },
  "variables": {
    "two_dmn_rec_var": {
      "shape": ["time", "lev"],
      "type": "float",
      "data": [[1.0, 2.0, 3.0], [1.0, 2.10, 3.0], [1.0,
2.20, 3.0], [1.0, 2.30, 3.0], [1.0, 2.40, 3.0], [1.0, 2.50,
3.0], [1.0, 2.60, 3.0], [1.0, 2.70, 3.0], [1.0, 2.80, 3.0],
[1.0, 2.90, 3.0]]
    }
  }
}
```

Multidimensional Arrays II

```
> ncks -C -H --jsn_fmt=4 -v two_dmn_rec_var in.nc
{
  "dimensions": {
    "lev": 3,
    "time": 10
  },
  "variables": {
    "two_dmn_rec_var": {
      "shape": ["time", "lev"],
      "type": "float",
      "data": [1.0, 2.10, 3.0, 1.0, 2.20, 3.0, 1.0, 2.30,
3.0, 1.0, 2.40, 3.0, 1.0, 2.50, 3.0, 1.0, 2.60, 3.0, 1.0,
2.70, 3.0, 1.0, 2.80, 3.0, 1.0, 2.90, 3.0, 1.0, 2.90, 3.0]
    }
  }
}
```

Enumerated Types

```
> zender@firn:~$ ncks --json enum.nc
{
  "types": {
    "enum_ubyte_t": [ "Clear":0, "Cumulonimbus":1, "Stratus":2,
"Missing":128 ],
  }
  "dimensions": {
    "lon": 4
  },
  "variables": {
    "cld_flg": {
      "shape": ["lon"],
      "type": "enum_ubyte_t",
      "attributes": {
        "_FillValue": Missing
      },
      "data": ["Stratus", "Missing", "Cumulonimbus", "Clear"]
    }
  }
}
```

Variable Length Arrays

```
> zender@firn:~$ ncks --json vlen.nc
{
  "types": {
    "int(*)" : "vlen_int_t",
  },
  "dimensions": {
    "lat": 2,
  },
  "variables": {
    "vlen_int_1D": {
      "shape": ["lat"],
      "type": "vlen_int_t",
      "attributes": {
        "_FillValue": [-999]
      },
      "data": [[17, 18, 19], [1, 2, 3, 4, 5, 6, 7, -2147483647, 9,
-2147483647]]
    }
  }
}
```

Groups

```
> zender@firn:~$ ncks --json grp.nc
{
  "groups": {
    "g1": {
      "variables": {
        "g1v1": {
          "type": "int"
          "data": 1
        }
      }
    },
    "g2": {
      "variables": {
        "g2v1": {
          "type": "int"
          "data": 2
        }
      }
    }
  }
}
```

NCO-JSON Adoption

1. NCO
2. ERDDAP
3. CF-JSON
4. STAR-JSON

Strengths & Weaknesses of ASCII netCDF Formats

	Strengths	Weaknesses
CDL	Legible, Complete	Niche
NcML	XML, THREDDS, OPeNDAP	netCDF4 gaps Verbose
JSON	JSON, Legible, (OPeNDAP)	???

Supplementary Slides