Observations and Measurements as a basis for semantic reconciliation between GRIB and netCDF

... and some other ideas.

Jeremy Tandy
24th September 2014
Problem statement: interoperability

interoperable

Pronunciation: /ɪntəˈprəpərəbəl/ 

Adjective 

(of computer systems or software) able to exchange and make use of information

... requires both structure (syntax) and meaning (semantics)
GRIB (WMO No. 306 Vol I.2 FM92-XIV GRIB) is ...

- a message format for regulated products ...
- comprised of sections:
  - data payload
  - serialization details (data-type, precision, packing, compression & ‘bitmap’ masks) ... syntax
  - content description (grid and product definitions) and fixed metadata (originating centre, reference time etc.) ... semantics
- table driven – governed sets of code-tables and templates; the GRIB ‘vocabulary’ (more semantics)
did anyone mention local tables?

• (section 2)
• designed for local use; bilateral exchange etc.
• often inaccessible beyond the original participants
• renders the GRIB message useless ... 
• no tables: no decode!
aside: publishing code-tables for reuse

http://codes.wmo.int/grib2/codeflag/4.2/0-0-0
NetCDF (Network Common Data Form) is ...

- an API for storing and retrieving data in form of arrays
- described with a **syntax of dimensions, variables and attributes**
- (plus **groups** in the NetCDF-enhanced model)
- HDF5 “under the hood” ... but those details are abstracted away by the API
NetCDF is ... just a ‘container’?

- out of the box, NetCDF can describe any array data ...
- data publishers are unconstrained in their choice of dimensions, variables, attributes and groups
- need conventions to supply semantics – so that consumers can understand the data too!

source: http://cfconventions.org/
Closing the gap #1

Publish data using well governed, publicly accessible semantics.
ISO 19156:2011 ‘Geographic information – Observations and measurements’ provides a metamodel for describing the context required to interpret the results of an observation.

OM_Observation:
- an EVENT whose
- RESULT is an estimate of a value of some
- PROPERTY of some
- THING obtained using a specified
- PROCEDURE ...

© Crown copyright
CF (CSML), CDM and O&M convergence

WMO and O&M convergence

- WMO tasked by ICAO to deliver GML application schema for aeronautical meteorology data exchange
- WMO (TT-AvXML) adopt model-driven approach based on O&M; new code forms include
  - FM 202-15 Ext. METCE-XML
  - FM 205-15 Ext. IWXXM-XML

O&M part of WMO Technical Regulation
Closing the gap #2

Agree that O&M provides the common top-level semantics for our community and determine how to map to that model.
Heterogeneous GRIB templates

*Product definition template 4.0 - analysis or forecast at a horizontal level or in a horizontal layer at a point in time*

<table>
<thead>
<tr>
<th>Octet No.</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Parameter category (see Code table 4.1)</td>
</tr>
<tr>
<td>11</td>
<td>Parameter number (see Code table 4.2)</td>
</tr>
<tr>
<td>12</td>
<td>Type of generating process (see Code table 4.3)</td>
</tr>
<tr>
<td>13</td>
<td>Background generating process identifier (defined by originating centre)</td>
</tr>
<tr>
<td>14</td>
<td>Analysis or forecast generating process identifier (defined by originating centre)</td>
</tr>
<tr>
<td>15–16</td>
<td>Hours of observational data cutoff after reference time (see Note)</td>
</tr>
<tr>
<td>17</td>
<td>Minutes of observational data cutoff after reference time</td>
</tr>
<tr>
<td>18</td>
<td>Indicator of unit of time range (see Code table 4.4)</td>
</tr>
<tr>
<td>19–22</td>
<td>Forecast time in units defined by octet 18</td>
</tr>
<tr>
<td>23</td>
<td>Type of first fixed surface (see Code table 4.5)</td>
</tr>
<tr>
<td>24</td>
<td>Scale factor of first fixed surface</td>
</tr>
<tr>
<td>25–28</td>
<td>Scaled value of first fixed surface</td>
</tr>
<tr>
<td>29</td>
<td>Type of second fixed surface (see Code table 4.5)</td>
</tr>
<tr>
<td>30</td>
<td>Scale factor of second fixed surface</td>
</tr>
<tr>
<td>31–34</td>
<td>Scaled value of second fixed surface</td>
</tr>
</tbody>
</table>

... there is no-size fits all mapping.
Closing the gap #3

Map GRIB templates to O&M on a case by case basis*.

*(and incorporate the O&M model into GRIB edition 3 to make the mapping easier)
Because GRIB templates are designed to meet a specific purpose, experts can select the information content on an ‘as-needed’ basis ... and hopefully map that onto well-known content models.
Supporting an ecosystem of conventions

CF is great – but not the whole story ...

inclusion of elements from other standards and conventions in NetCDF alongside CF is difficult:

• validation tools throw exceptions
• reliance on NetCDF-classic model (NetCDF3) prevents use of **groups**
i) namespace awareness in NetCDF

Because attributes and variables are scoped to a group; names clashes can be avoided and software tools can ignore unrecognized properties ...

- Discovery & usage metadata
- Probability density function
- Climate prediction model data
ii) including (complex) objects

Nesting of groups allows inclusion of complex objects...

```
<group name="citation">
  <attribute name="uuid" value="UUID"/>
  <attribute name="objectType" value="acdd:CI_Citation"/>
  <attribute name="title" value="Title of cited item"/>
  <attribute name="identifier" value="Identifier of cited item"/>
  <attribute name="edition" value="Edition of cited item"/>
  <group name="date">
    <attribute name="date" value="Date associated with cited item"/>
    <attribute name="dateType" value="Type of date associated with cited item"/>
  </group>
</group>

<group name="citedResponsibleParty">
  <attribute name="uuid" value="UUID"/>
  <attribute name="objectType" value="acdd:CI_ResponsibilityParty"/>
  <attribute name="Name of responsible individual"/>
  <attribute name="organisationName" value="Name of responsible organisation"/>
  <attribute name="electronicMailAddress" value="Email of responsible individual"/>
  <group name="onlineResource">
    <attribute name="uuid" value="UUID"/>
    <attribute name="objectType" value="acdd:CI_OnlineResource"/>
    <attribute name="protocol" value="http"/>
    <attribute name="linkage" value="http://earthdata.nasa.gov"/>
    <attribute name="applicationProfile" value="Web Browser"/>
    <attribute name="name" value="EOSDIS - Earth Data Website"/>
    <attribute name="description" value="Access to data and information"/>
    <attribute name="function" value="information"/>
  </group>
</group>
```
Closing the gap #4

Update* the CF convention and related tools to support the NetCDF-enhanced model and namespace awareness using groups.

*(perhaps establishing “CF2” as a superset of CF is more appropriate than an update?)

Note there is an active discussion thread “CF Conventions and netCDF-4 enhanced model”
Summary: gaps to close ...

1. Publish data using well governed, publically accessible semantics.

2. Agree that O&M provides the common top-level semantics for our community and determine how to map to that model.

3. Map GRIB templates to O&M on a case by case basis.

4. Update the CF convention and related tools to support the NetCDF-enhanced model and namespace awareness using groups.
Thank you