GRIB Background & OGC Foreground

Some lessons learnt, some suggestions for the future

Chris Little, ECMWF, 25 Sept 2014

Why me?

OGC

- Co-Chair of Met Ocean Domain WG
- Co-Chair of Temporal Domain WG
- Establishing an OGC Vertical CRS WG

WMO

- Regional Association VI Co-Rapporteur for Information Systems & Services (with Leonid Bezruk, Russia)
- WIS Focal Point for GISC Exeter

Been around a long time

(I am partly to blame for GRIB, BUFR & CREX)

Contents

This presentation covers the following areas

- A Bit of GRIB (& NetCDF) History
- Some Principles
- Some OGC ideas
- Questions and answers

The History Bits

GRIB's parents

- WMO GRID
- US NWS H-Code
- ECMWF Internal format
- UKMO PP/Fieldfiles/UM file format
-
- IOC GF3 (BUFR really)
-
- etc

Context then (late 1980s)

- Octet/Byte orientated formats becoming inadequate
- Different hardware formats (IBM, DEC, Cray, CDC, etc)
- Telecoms costs significant
- CPU for compression/decompression to save I/O & CPU
- 16 bit PCs appearing
- Memory Access on 8xn boundaries faster
- Table driven software emerging, <256 entries not enough
- Binary telecoms protocols establishing (FTP/TCP/IP)

Some Principles

Some Attributes

- GRID:
  - Character not binary
  - Human readable (some character encodings not readable)
- Binary gives compression:
  - ASCII “0123456789” -> 70 or 80 bits
  - BCD “0000 0001 0010 0011 0100 0101 0110 0111 1000 1001” -> 40 bits, ~x2 compression
- Scaled +ve integers as bits gives more compaction
  - Avoids hardware number representation differences
  - Avoids conversion between machines
  - CPU efficient

Why gridded data?

- Only store data once, if possible
  - Definitely not for every grid point
  - Find location by lightweight counting, not heavyweight calculations
  - Some calculations avoided by storing (x,y) with every grid point
- Could be Binary or Character
- Local Compaction vs Global Compression
  - I.e. counting across compacted form finds individual values
  - Global decompression needed to find individual values

GRIB 2 that never was

- WMO voted to develop GRIB2
- Outvoted:
  - NetCDF + Compaction/compression + WMO Tables -> GRIB2
- No money to fund UCAR

Mistakes and Regrets

Never did GRIB (or BUFR) API or object orientated description

- Constrained by time, other work
- Constrained by WMO Regulatory text format

Envisaged more shared components between GRIB & BUFR

Envisaged BUFR & CREX having exactly same objects & API

- ITU/ISO fashion for functional standard & various encodings/bindings
- E.g. ISO CGM, ITU ASN.1 BER

GRIB never did:

- Run Length Encoding
- Polar orbit swathe
- Cross sections, Hovmöller diagrams

Mistakes and Regrets (BUFR)

IOC GF3 format allows any parameter to be a coordinate, by setting a flag

- Could have partitioned Table B similarly

Never re-factored BUFR soon enough

- BUFR-lite <-> CREX
- Operators too complex
- Operators and Attributes modal, not context-free scope
- No API

Other progress and problems

- Introduced fast track for non-contentious changes
- Introduced IETF interoperable implementations criterion
  - Stopped upper management tweaking
- Table stuffing:
  - Get the last vacant slots for our NMS specific requirements
  - Break the underlying conceptual model
  - Break the spirit of the regulations:
    - “All our parameters have a range of ±2**31”

Some OGC Ideas

For both GRIB and NetCDF

Temporal stuff

OGC clarifying geospatial use of Time:

- In a time stamp:
  - Is value 60 minutes valid? 2014-09-26T00:60:00.0 ?
  - Is value 60 seconds valid? 2014-09-26T00:00:60.0 ?
  - Is leap second valid? 2012-06-30T23:59:60.0 ?
- Calculating durations:
  - Is (2012-07-01T00:00:00.0) – (2012-06-30T23:59:00.0) = 60 or 61?
- A Calendar is not a CRS is not a Calendar

OGC established registry for Temporal CRS definitions

Temporal Regimes:

1. Events, partially ordered, Allen's 13 logical operators, no clocks, no instants, no durations
   E.g. ice cores, archaeology, tree rings, king lists

2. Clocks, timescale, fixed precision ticks, integer arithmetic, integer instants, integer durations, epoch/datum
   E.g. TAI Atomic time, pendulum, earth rotation

3. CRSs, real number line, normal arithmetic, +ve / –ve directions, instants, durations, epoch/datum
   E.g. Julian Days, Unix Milliseconds

4. Calendar, real number line, abnormal arithmetic, algorithms
   E.g. UTC Gregorian Calendar, Mayan

ISO 8601 is notation with various pieces of above

Vertical CRS ‘regimes’

Several types of Vertical CRS:

- Datums:
  - Fixed datum
  - Time and space varying datum
- Axis and unit
  - Simple
  - Combinations
  - Non-monotonic

Conceptual model needed. Collecting examples now

GRIB is fuzzy between parameter, cell representation (Z & XY)

Parameters

- Not clear what is a ‘parameter’:
  - CF/GRIB/BUFR/UKMO Stash/etc
  - Parameter values vs Coordinate values
  - Representativeness (spatial and temporal average vs instantaneous point value)

Questions & answers

Discussion, Any Conclusion?