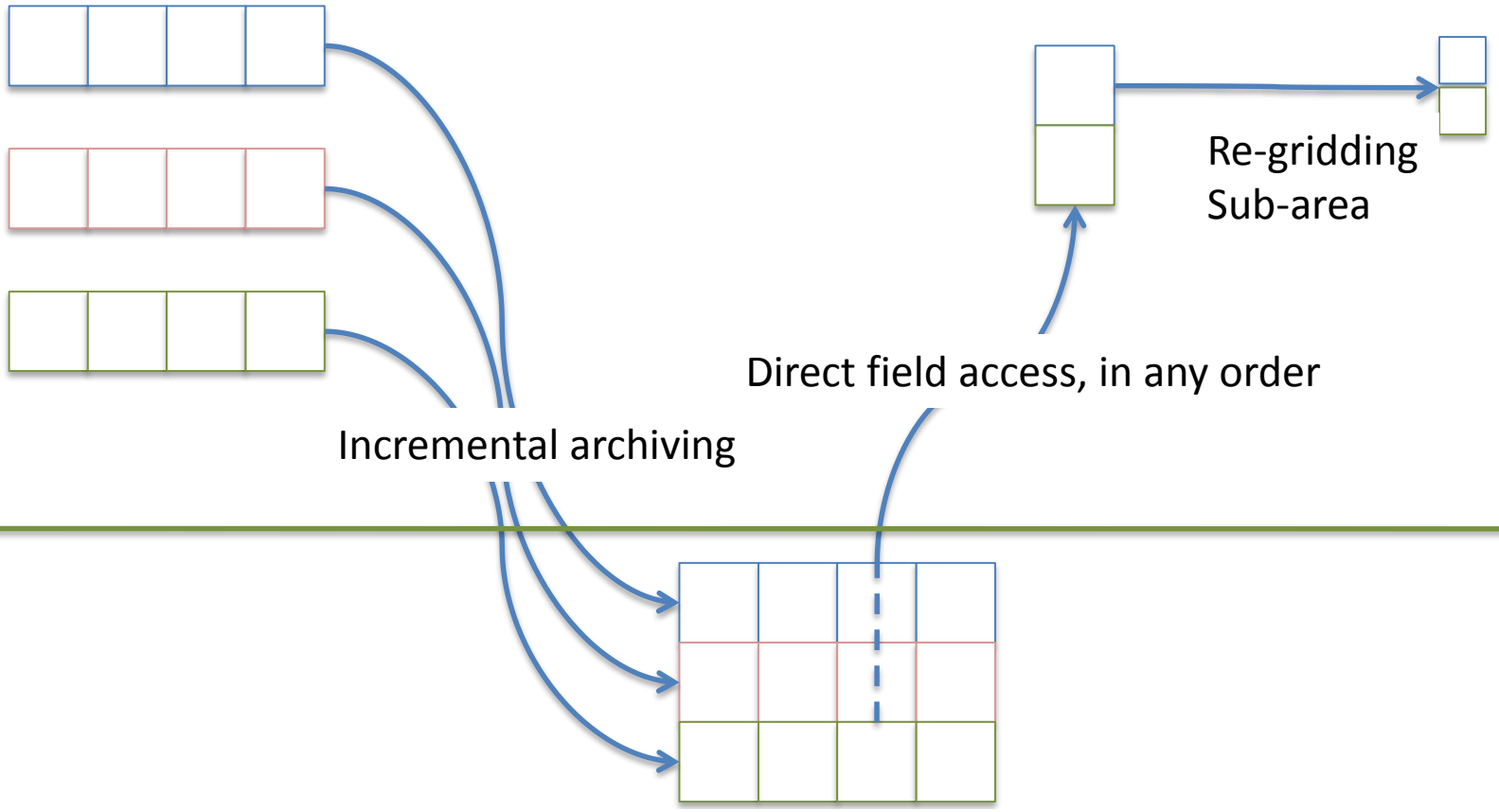
NetCDF support in MARS

Manuel Fuentes

What services does MARS offer?



How is it done

- Currently MARS handles 2D fields in GRIB
- GRIB data are records not files
- MARS scans archived files and extracts metadata from GRIB headers
- GRIB are reorganised into larger files, to minimise the total number of files and collocate related fields to speed up retrievals
- MARS keeps an index that tracks where each GRIB field is (file, offset, length)
- On retrievals, MARS find the required fields, reads them from tape, and re-assembles them according to the user's request
- MARS can also perform re-gridding or sub-area extraction

MARS in NetCDF

- We should provide the same services for NetCDF
  - Incremental archive
  - Data collocation
  - User can select any 2D fields from the archive and have them delivered in a single file
  - Re-gridding and sub-area extraction
- Problem:
  - NetCDF is a file format, not a record format
  - Original files contain multi-dimensional variables (often > 2 dimensions)
  - One cannot extract a 2D field from a NetCDF file directly from tape

Solution considered

- NetCDF files to be archived are transparently split into individual NetCDF files containing a single 2D field
- Resulting NetCDF files are annotated with MARS specific information, using NetCDF file ‘Variable’ attributes
- These attributes are used by MARS to index the NetCDF files, and treat them as simple binary records
- On retrieval, those records will be assembled in a single NetCDF3/4 file to be delivered to the user
- Re-gridding and sub-area will be implemented at a latter stage (new interpolation software currently being developed)
- The delivered NetCDF files must be CF compliant (with valid CF “standard name” attributes attached to the variables), and ideally as close as possible to the CMIP5/OGC standards. Non-CF compliant files will be rejected.