



WP-MIP FORECAST ARCHIVE

2ND WP-MIP PROJECT MEETING
1500-1600 UTC 31 JULY 2025



AGENDA

PROJECT UPDATE

FORECAST DATA ARCHIVE

NEXT STEPS



PROJECT UPDATE

Progress has been made on a number of topics:

- Whitepaper is largely “frozen”
- Project membership is strong (>70)
- Expecting forecast contributions from 17 institutions on 6 continents
- WP-MIP is a key project in the WIPPS AI Roadmap, which underpins a draft Council resolution
- Project announcements made in the ESMO Newsletter and submitted to WIPPS Newsletter



FORECAST DATA ARCHIVE

Original Plan

- Forecasts delivered by contributing centres to ECMWF in grib2 format via ftp
- Data restructured for internal MARS requirements
- Users access data through MARS interface as for ERA5, TIGGE, S2S, etc
- Benefits of MARS: a very well-defined, flexible and familiar user interface

The Bad News

- The ECMWF Council voted to not host WP-MIP data in early July
- Decision appears to be based on politics rather than scientific value
- Efforts to negotiate for MARS space have failed



CONTINGENCY PLAN

Environment and Climate Change Canada (ECCC) has offered to host the WP-MIP dataset

- A project account has been created and sufficient space (50TB) allocated to begin building the dataset
- Additional space can be added as the project dataset expands
- Data access for diagnostics and evaluation will be through ftp/scp rather than the MARS interface
- Only file retrieval (no subsetting or computing) is supported



DATASET STRATEGY

Data format will continue to be **grib2**:

- Simplifies operational evaluation by following existing exchange workflows
- Supports contributors who have already prepared their conversion tools
- Protocol evaluation tools are in progress

Contributor submission files **will remain unchanged from the whitepaper protocol**: file manipulation will be done when we add the contribution to the WP-MIP database

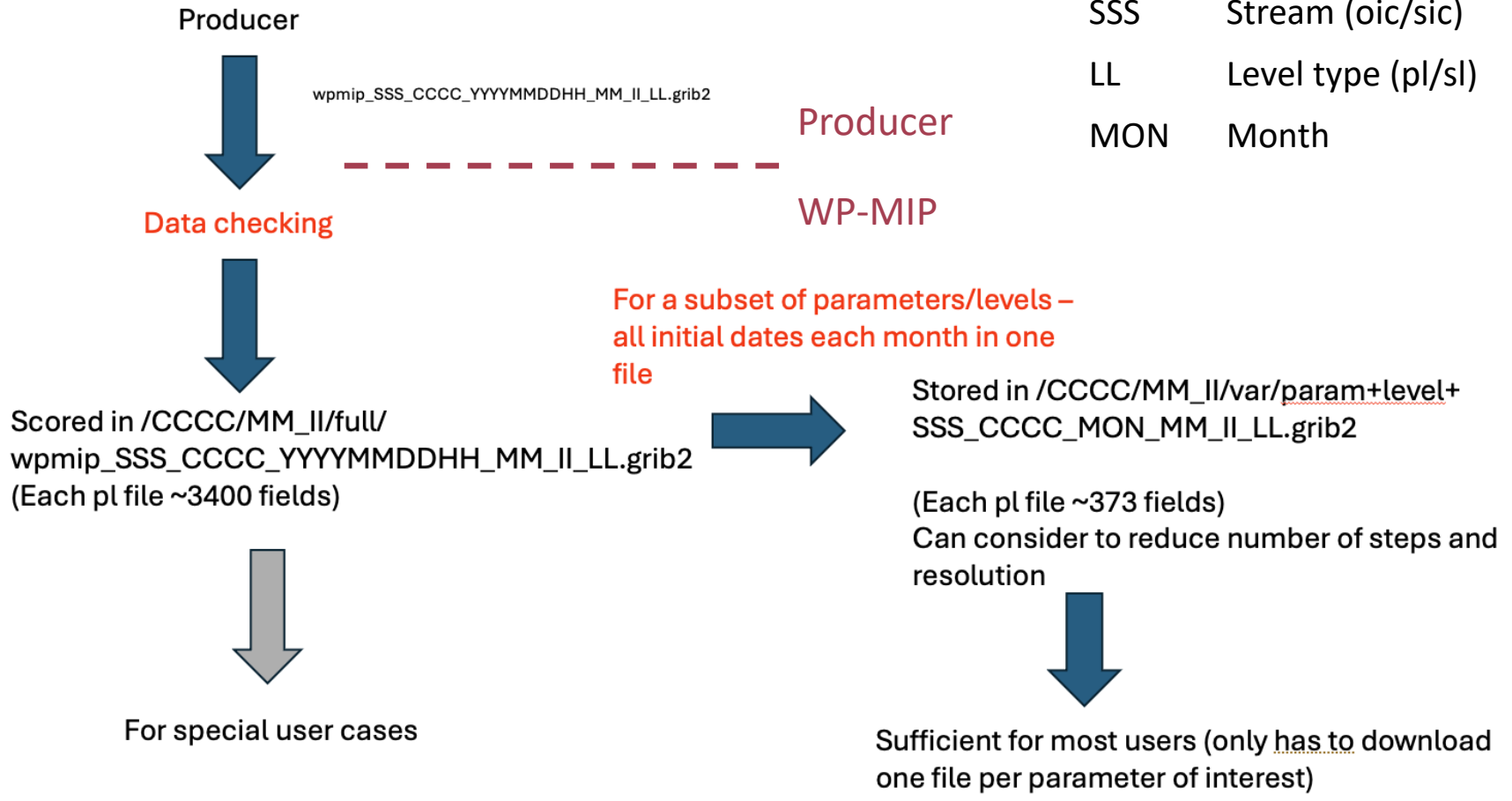
Finite storage space prompts us to propose two minor protocol changes:

1. Provide only 10, 50 and 100 hPa in the stratosphere (drop 1, 2, 3, 5, 20)
2. Provide only 24-hourly outputs after 240h for SP2 integrations (instead of 6-hourly)

DATASET LAYOUT

CCCC	Centre acronym
MM	Model type (pm/ai/hy)
II	Iteration number (from 0)
SSS	Stream (oic/sic)
LL	Level type (pl/sl)
MON	Month

Full-file downloads means that we should be strategic in our database design so that users don't have to download huge data volumes just to extract a small subset.



PROTOTYPE CHECKING

We have a preliminary file checking tool ready (at ECMWF) for pressure-level prototype files.

Checking tool needs to be extended to single-level files.

We are building an issue tracking system to help follow iterations with contributors and track progress.

Example checker results.

- CMC files:
 - FAIL: scanningMode(64) == 0
 - FAIL: latitudeOfFirstGridPoint(-90000000) == 90000000
 - FAIL: latitudeOfLastGridPoint(90000000) == -90000000
- The JMA files
 1. First the AI model file check. This is the only one.
 - FAIL: versionNumberOfGribLocalTables(1) == 0Again, I am not sure how important it is to get this right...
 2. Then physics model check
 - FAIL: Wpmip Basic Checks
 - FAIL: productionStatusOfProcessedData(2) in [16, 17]
 - FAIL: backgroundProcess(2) == 1
 - FAIL: generatingProcessIdentifier(255) == 3
 - FAIL: gribMasterTablesVersionNumber(2) >= 4
 - FAIL: versionNumberOfGribLocalTables(1) == 0



NEXT STEPS

- Finalize grib checking tools and install on ECCC servers
 - Retrieve ssh public keys from contributors for server access
 - **Receive and check (iterate) prototype files from contributors:** a full set of forecast files for a single date
 - Investigate long-term alternatives with enhanced functionality (e.g. NERC JASMIN platform) possibly for eventual database transfer
 - Receive short (half-page) model descriptions from contributors for documentation and project overview paper
- 