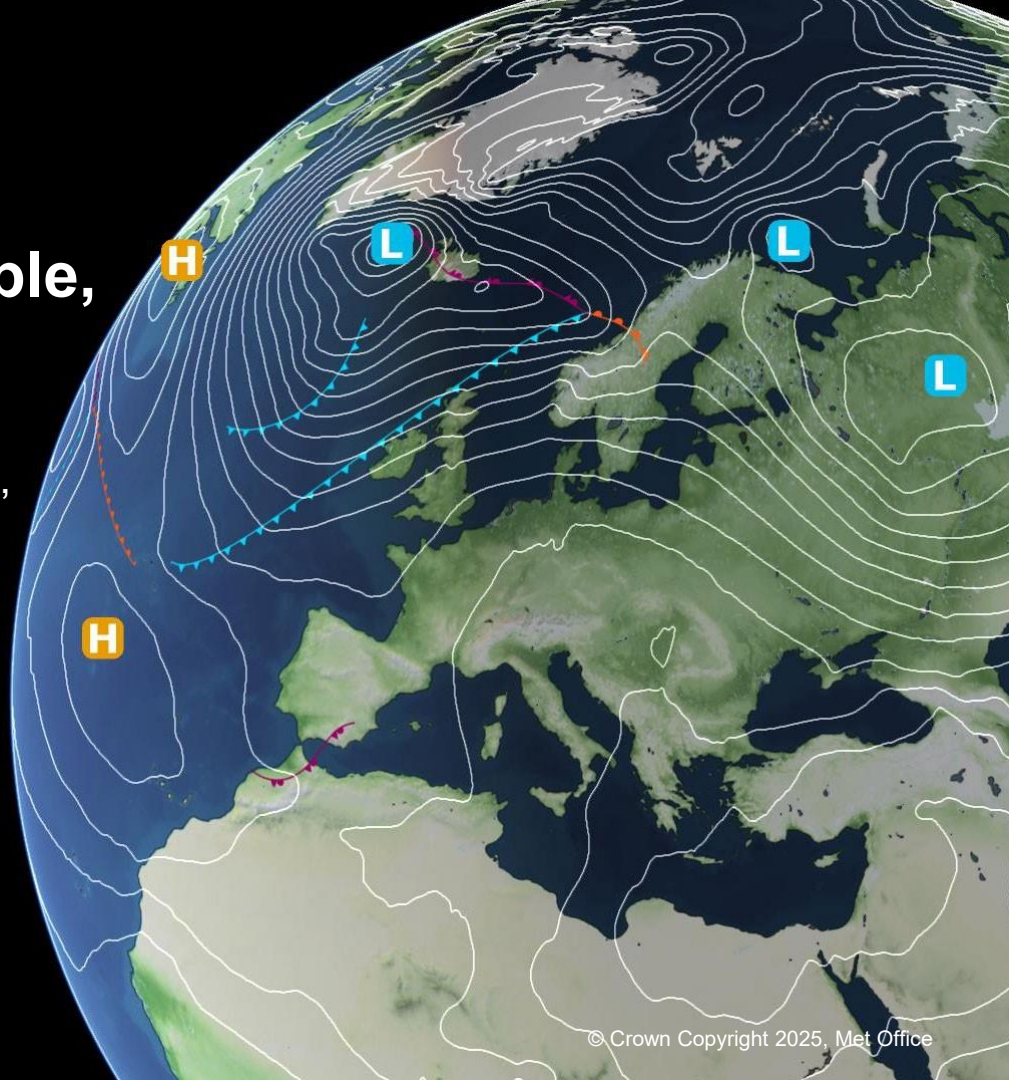
Met Office
Hadley Centre

Climate Data Storage – simple, consistent and traceable

Dan Hollis, Mike Kendon, Stephen Packman, Emily Carlisle and Amy Doherty

15th EUMETNET Data Management Workshop

Oslo, 4-6 November 2025



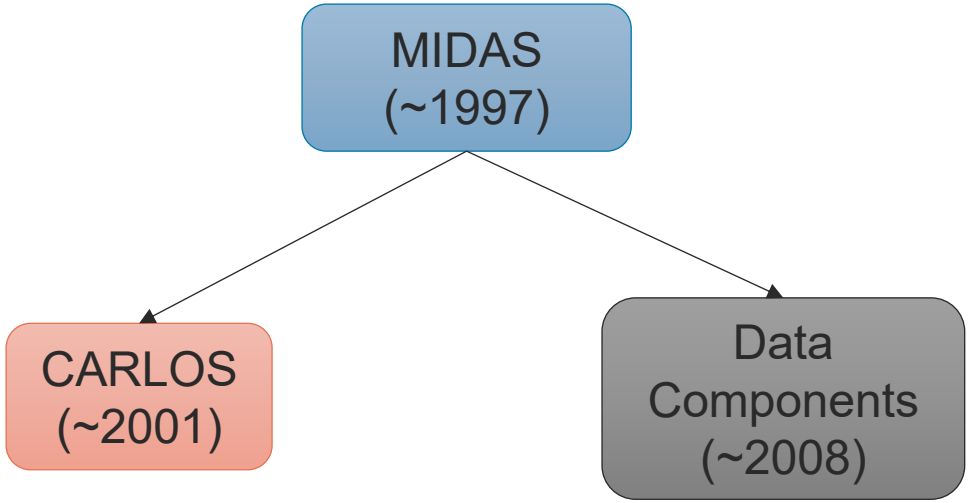
Met Office
Hadley Centre

Background

- The National Climate Information Centre (NCIC) is responsible for monitoring the climate of the UK
- 20+ years ago we created an Oracle database (CARLOS) of monthly climate statistics at stations
- This database is populated by legacy FORTRAN code
- The platform on which the software runs will be decommissioned next year
- We are now designing and building a replacement system

Met Office
Hadley Centre

What we currently do

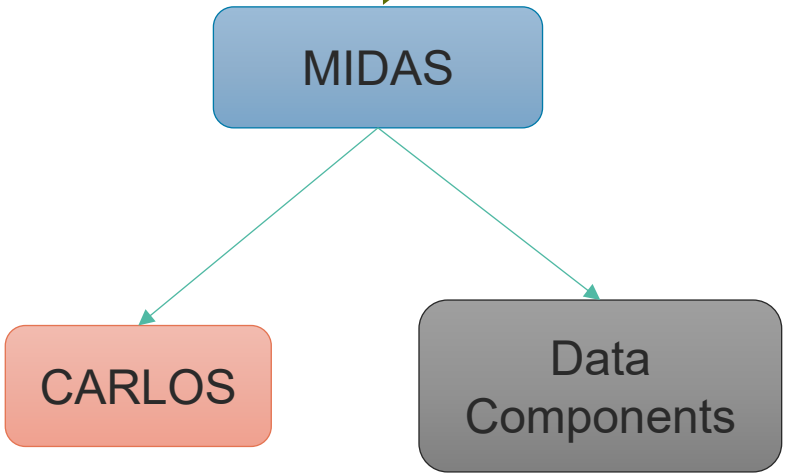


MIDAS = Met Office Integrated Data Archive System
CARLOS = Climatological And Related Library Of Statistics

Met Office
Hadley Centre

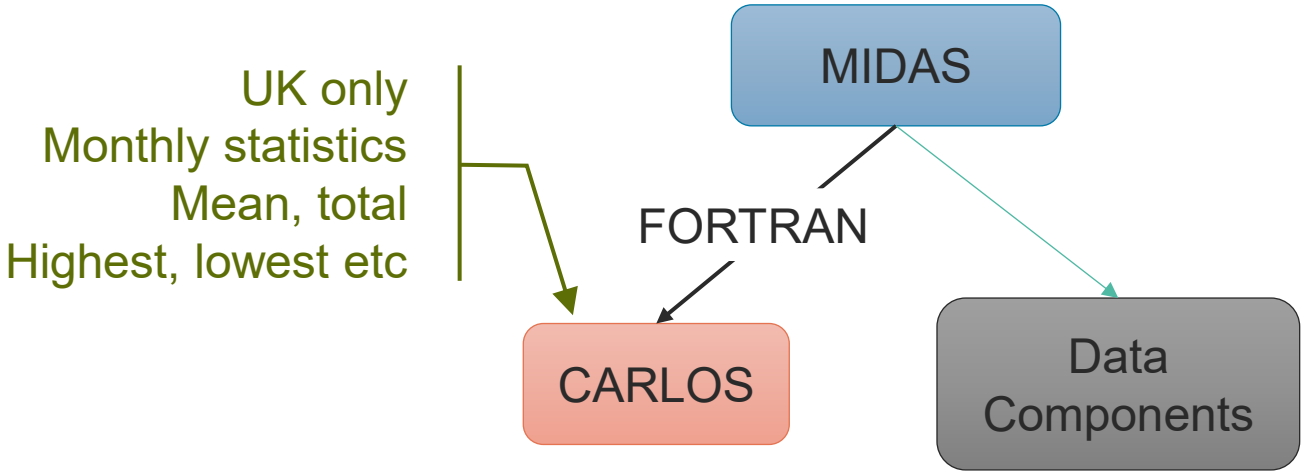
What we currently do

Hourly, daily and monthly data
Global
Land, marine and upper air
Raw observations
Quality controlled observations



Met Office
Hadley Centre

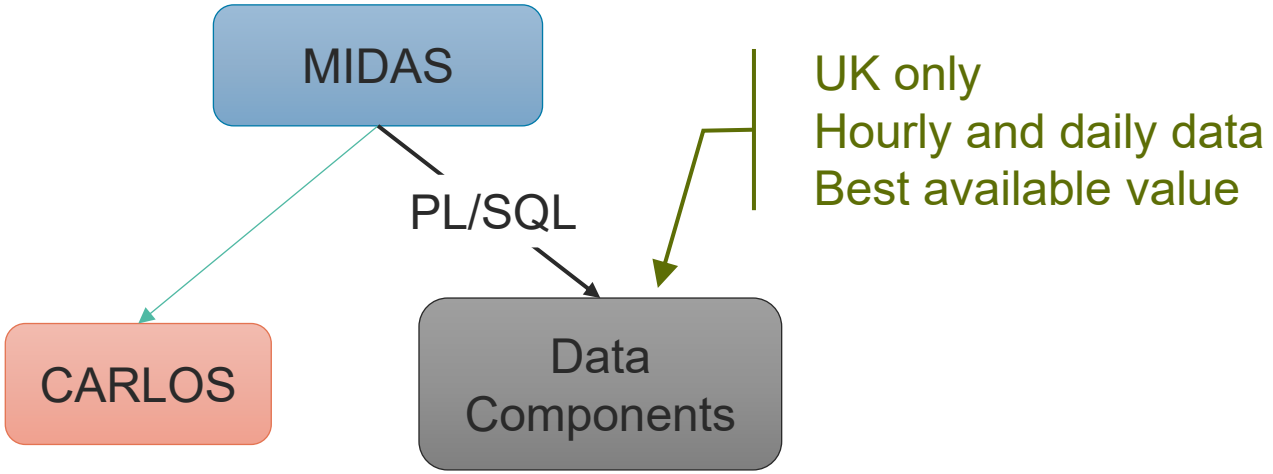
What we currently do



Met Office
Hadley Centre

What we currently do

Different observing practices
Multiple rain gauges
Multiple database tables
Multiple station identifiers
Complex metadata



Met Office
Hadley Centre

Challenges

- History (legacy systems, diminishing knowledge, evolving IT)
- Resources (staff, infrastructure)
- Consistency (with other databases)
- Dependent systems (derived datasets, internal/external web sites, customer services)
- Different data origins (calculated values, rescued data)
- Multiple use cases (data storage, user applications)
- Data licensing

Met Office
Hadley Centre

Requirements

- Monthly statistics
- Means and totals
- Counts of days
- Highest and lowest daily values
- Metadata

- Simple and modular
- Consistent
- Timely
- Accessible
- Scalable
- Traceable
- Portable
- Well-documented

Met Office
Hadley Centre

Proposed design

- **New starting point**
- Data Components
- (instead of MIDAS)

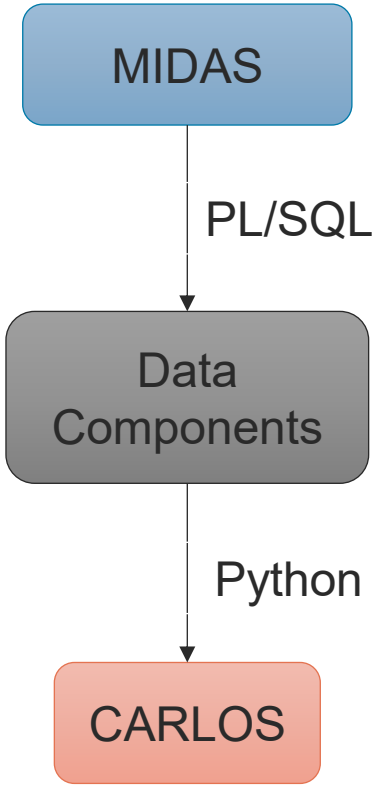
- **New software**
- Python
- (instead of FORTRAN)

- **New table design**
- Thin table
- (instead of a wide table)

Met Office
Hadley Centre

Progress

- Design decisions
  - Revised flow diagram
  - Programming language
  - New table structure
- Coding
  - Software environment / installation process
  - Configuration (YAML)
  - Calculation of simpler variables



Met Office
Hadley Centre

Table structure – old vs new

SRC_ID	MET_ELEM_MISS_HOUR_CNT
MET_DOMAIN_NAME	MET_ELEM_QUAL_ID
OB_DATE	MET_ELEM_SBST_CODE
MET_ELEMENT_ID	MET_ELEM_SBST_HOUR_CNT
MET_ELEM_AMT	MET_ELEM_SD_VAL
MET_ELEM_DAY_CNT	MO_SFC_ELEM_PRTN_ID
MET_ELEM_HOUR_CNT	REC_ST_IND
MET_ELEM_MAX_VAL	VERSION_NUM
MET_ELEM_MAX_VAL_DAY_NUM	LAST_UPDATE_TIME
MET_ELEM_MEAN_VAL	
MET_ELEM_MIN_VAL	
MET_ELEM_MIN_VAL_DAY_NUM	
MET_ELEM_OCCR_PRC	

id
id_type
report_type
result_time
phenomenon_duration
parameter_name
statistic
data_value
data_units
data_qc
data_version
stamp_time

Met Office
Hadley Centre

Table structure – old vs new

SRC_ID
MET_DOMAIN_NAME
OB_DATE
MET_ELEMENT_ID
MET_ELEM_AMT
MET_ELEM_DAY_CNT
MET_ELEM_HOUR_CNT
MET_ELEM_MAX_VAL
MET_ELEM_MAX_VAL_DAY_NUM
MET_ELEM_MEAN_VAL
MET_ELEM_MIN_VAL
MET_ELEM_MIN_VAL_DAY_NUM
MET_ELEM_OCCR_PRC
MET_ELEM_MISS_HOUR_CNT
MET_ELEM_QUAL_ID
MET_ELEM_SBST_CODE
MET_ELEM_SBST_HOUR_CNT
MET_ELEM_SD_VAL
MO_SFC_ELEM_PRTN_ID
REC_ST_IND
VERSION_NUM
LAST_UPDATE_TIME

id
id_type
report_type
result_time
phenomenon_duration
parameter_name
statistic
data_value
data_units
data_qc
data_version
stamp_time

Met Office
Hadley Centre

Table structure – old vs new

SRC_ID
MET_DOMAIN_NAME
OB_DATE
MET_ELEMENT_ID
MET_ELEM_AMT
MET_ELEM_DAY_CNT
MET_ELEM_HOUR_CNT
MET_ELEM_MAX_VAL
MET_ELEM_MAX_VAL_DAY_NUM
MET_ELEM_MEAN_VAL
MET_ELEM_MIN_VAL
MET_ELEM_MIN_VAL_DAY_NUM
MET_ELEM_OCCR_PRC

MET_ELEM_MISS_HOUR_CNT
MET_ELEM_QUAL_ID
MET_ELEM_SBST_CODE
MET_ELEM_SBST_HOUR_CNT
MET_ELEM_SD_VAL
MO_SFC_ELEM_PRTN_ID
REC_ST_IND
VERSION_NUM
LAST_UPDATE_TIME

id
id_type
report_type
result_time
phenomenon_duration
parameter_name
statistic
data_value
data_units
data_qc
data_version
stamp_time

Met Office
Hadley Centre

Configuration – example 1

```
35222:
    name: 24hr maximum air temperature
    mdc_id: 40027
    temporal_resolution: "daily"
    stats: ["mean", "max", "min"]
    units: "degC"
```

Met Office
Hadley Centre

Configuration – example 2

```
35241:
    name: Maximum air temperature >= 25C
    mdc_id: 40027
    temporal_resolution: "daily"
    filter: {mdc_value: [">=", 25]}
    stats: ["count"]
    units: "degC"
```

Met Office
Hadley Centre

Configuration – example 3

```
35219:
    name: Diurnal temperature range
    mdc_id: [40027,40000]
    temporal_resolution: "daily"
    func: "difference"
    stats: ["mean", "max", "min"]
    units: "degC"
```

Met Office
Hadley Centre

Next steps

- Coding up the more complex variables (sunshine, vapour pressure)
- Create and populate new database table
- Compare new and old tables (are the calculated stats the same?)
- Identify rescued data in the old database and migrate to new table
- Implement automated update process (keep in step with MIDAS and Data Components)
- A new name!

Met Office
Hadley Centre

Summary

- Simple?
  - The design of the new system will be easier to work with (both the database itself and the software used to populate it).
- Consistent?
  - CARLOS will stay in step with MIDAS and Data Components, ensuring consistent data for all users.
- Traceable?
  - It will be easier to see which data are calculated and which data are rescued.