TIGGE and the EU Funded BRIDGE project

Baudouin Raoult

Head of Data and Services Section

ECMWF

ECMWF

The TIGGE core dataset

- **T**HORPEX **I**nteractive **G**rand **G**lobal **E**nsemble

- **Global ensemble forecasts to around 14 days generated routinely at different centres around the world**

- **Outputs collected in near real time and stored in a common format for access by the research community**

- **Easy access to long series of data is necessary for applications such as bias correction and the optimal combination of ensembles from different sources**

ECMWF

Phased implementation of the archive

● **Phase 1: multiple instances, low development effort**

**Phase 1
Archive
Centres**

ECMWF   **NCAR**   **CMA**
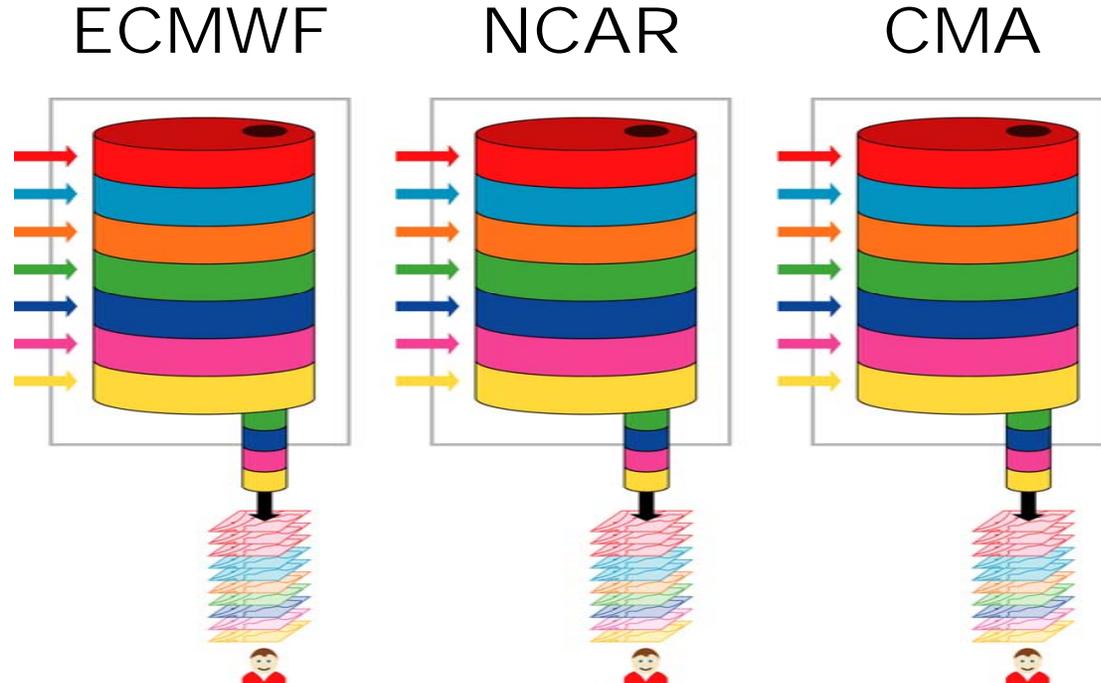
ECMWF

Building the TIGGE database

- **Three archive centres: CMA, NCAR and ECMWF**

- **Ten data providers:**

  - **Already sending data routinely: ECMWF, JMA (Japan), UK Met Office (UK), CMA (China), NCEP (USA), MSC (Canada), Météo-France (France), BOM (Australia),**

  - **Coming soon: CPTEC (Brazil), KMA (Korea)**

- **Exchanges using UNIDATA LDM, HTTP and FTP**

- **Operational since 1st of October 2006**

- **77 TB, growing by ~ 1 TB/week**

ECMWF

TIGGE Archive Centres and Data Providers

UKMO

CMC

ECMWF

NCAR

NCEP

MeteoFrance

CMA

JMA

KMA

CPTEC

BoM

- Archive Centre
- Current Data Provider
- Future Data Provider
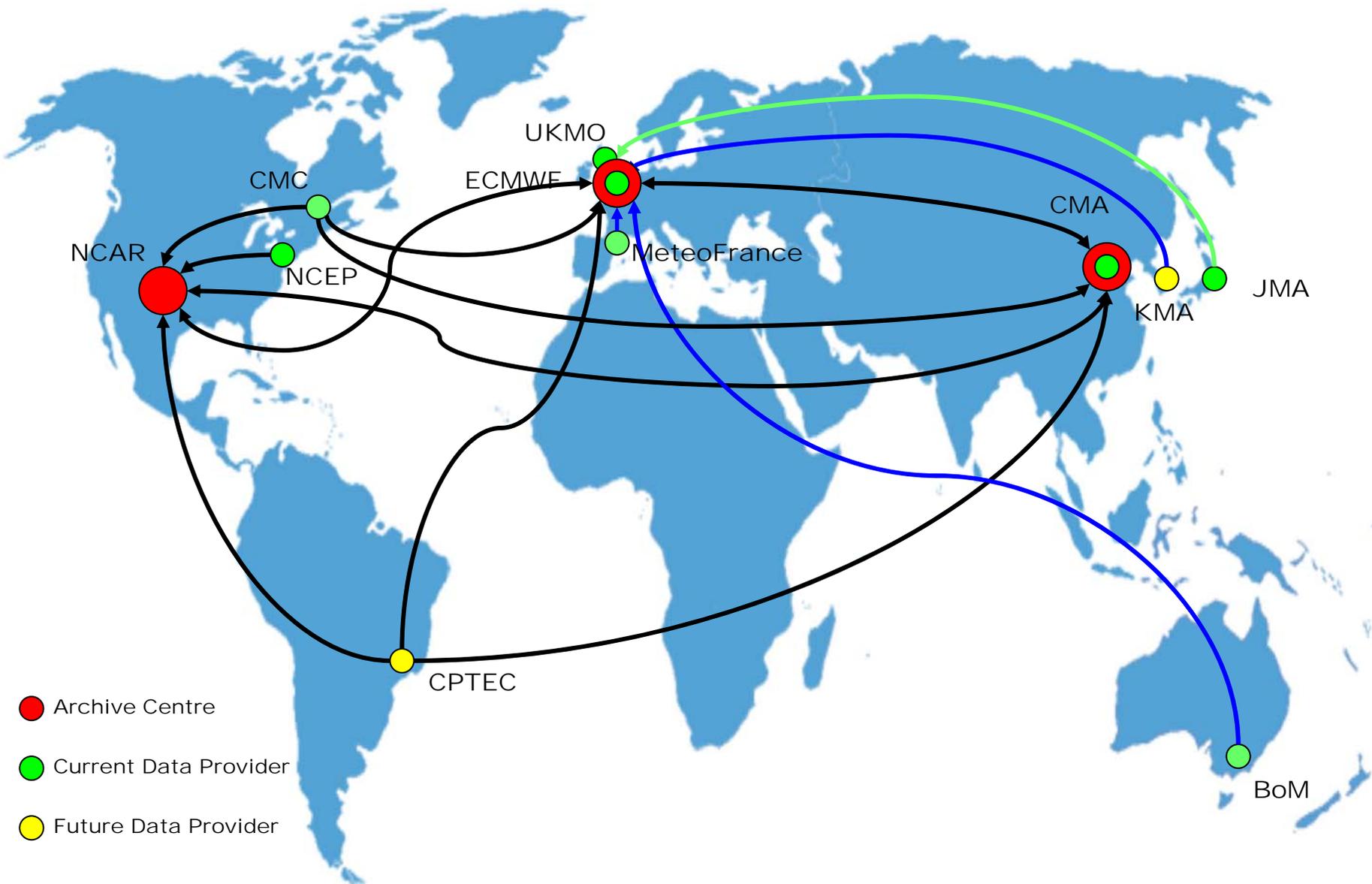
Strong governance

- **Precise definition of:**
  - **Which products: list of parameters, levels, steps, units,…**
  - **Which format: GRIB2**
  - **Which transport protocol: UNIDATA's LDM**
  - **Which naming convention: WMO file name convention**

- **Only exception: the grid and resolution**
  - **Choice of the data provider**
  - **Best possible model output**

- **Many tools and examples:**
  - **Sample dataset available**
  - **Various GRIB2 tools, "tigge_check" validator, …**
  - **Scripts that implement exchange protocol**

ECMWF

Using SMS to handle TIGGE flow
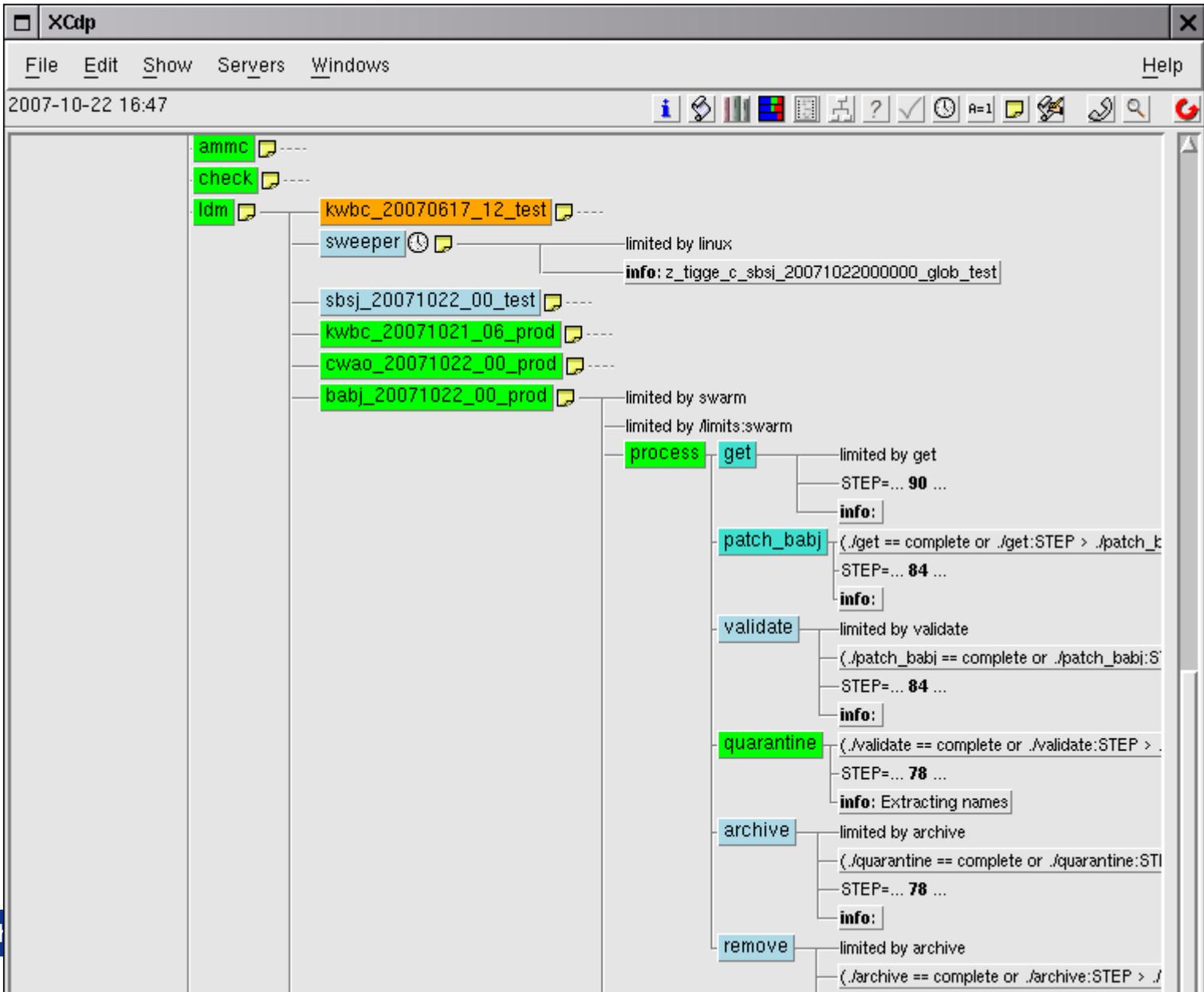
Quality assurance: homogeneity

- **Homogeneity is paramount for TIGGE to succeed**
  - The more consistent the archive the easier it will be to develop applications

- **There are three aspects to homogeneity:**
  - Common terminology (parameters names, file names,…)
  - Common data format (format, units, …)
  - Definition of an agreed list of products (Parameters, Steps, levels, …)

- **What is not homogeneous:**
  - Resolution
  - Base time (although most provider have a run a 12 UTC)
  - Forecast length
  - Number of ensemble

ECMWF

QA: Checking for homogeneity

Land sea mask (sfc), step 48, 20071107
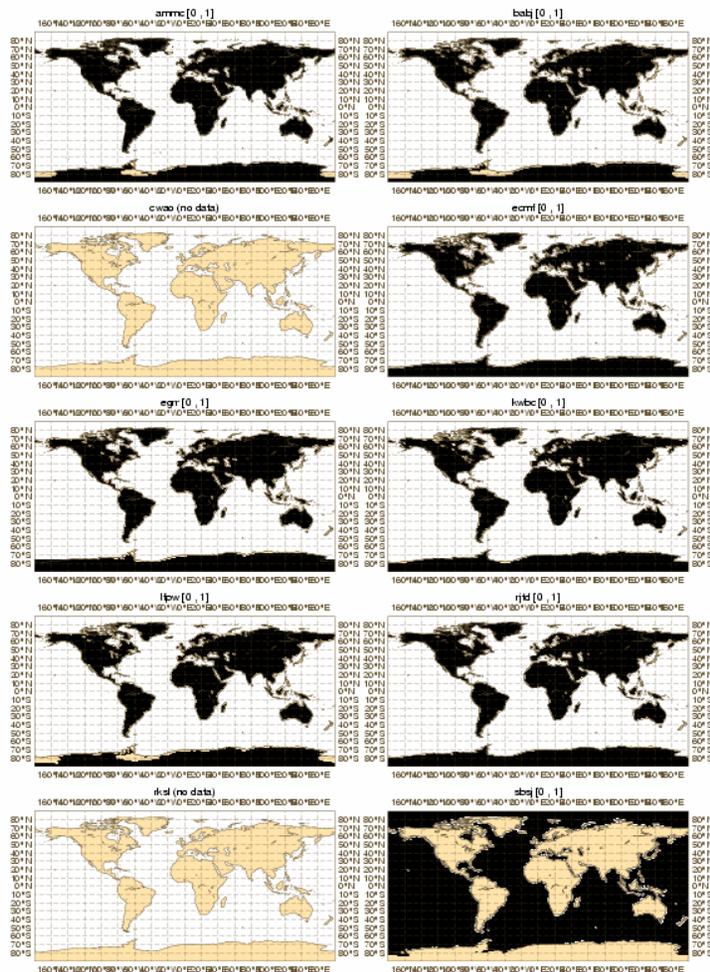
Time integrated surface net solar radiation (sfc), step 96, 20070608

ECMWF
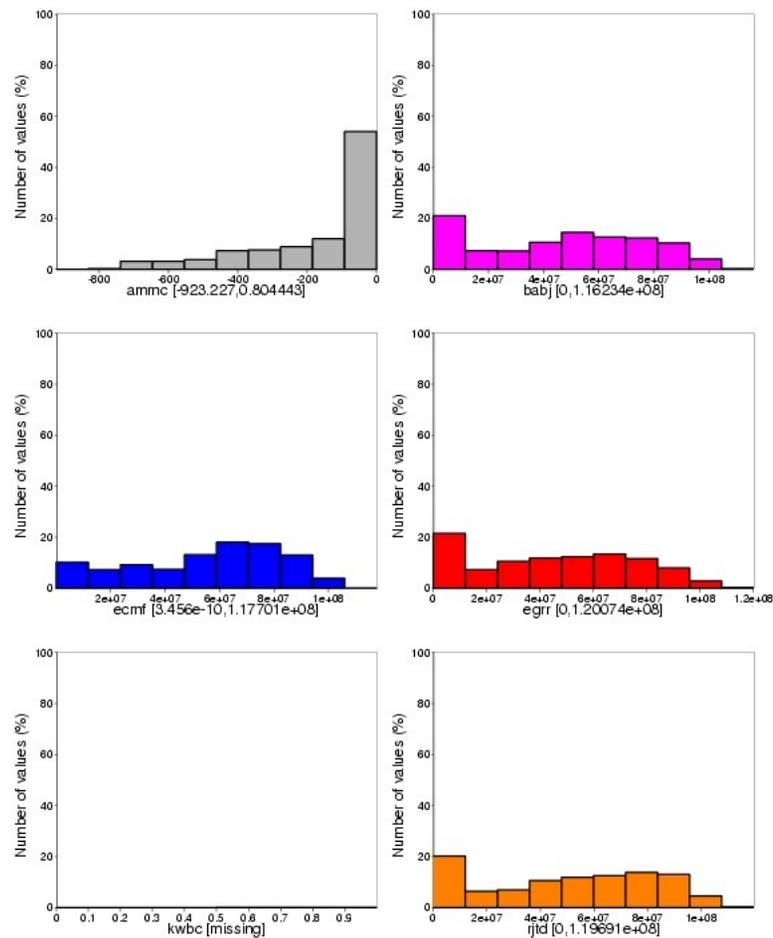
QA: Completeness

- **The objective is to have 100% complete datasets at the Archive Centres**

- **Completeness may not be achieved for two reasons:**

  - **The transfer of the data to the Archive Centre fails**

  - **Operational activities at a data provider are interrupted and back filling past runs is impractical**

- **Incomplete datasets are often very difficult to use**

- **Most of the current tools (e.g. epsgrams) used for ensemble forecasts assume a fixed number of members from day to day**

  - **These tools will have to be adapted**

ECMWF

QA: Checking completeness

Date												
												missing cycle
2007-06-11			97 missing fields									
2007-06-14										1302 missing fields	300 missing fields	
2007-06-17									missing cycle			missing cycle
2007-06-18									missing cycle	missing cycle	missing cycle	missing cycle
2007-06-19									missing cycle	missing cycle	missing cycle	missing cycle
2007-06-20									missing cycle	missing cycle	missing cycle	missing cycle
2007-06-26									missing cycle	missing cycle		
2007-07-17											1 missing field	2 missing fields
2007-08-13									missing cycle	missing cycle		
2007-09-03	starting 73986 fields	starting 73986 fields										
2007-09-28									missing cycle			
2007-10-01										missing cycle		
2007-10-03					starting 81837 fields	starting 81837 fields						
2007-10-09		615 new fields	615 new fields									missing cycle
2007-10-11												

Accessing TIGGE

- **Data is available to Research and Education after 48 hours**

  - **Self registration by agreeing to the terms and conditions**

- **Portals at CMA, NCAR and ECMWF.**

- **ECMWF portal offers:**

  - **Access to offline data**

  - **Aggregation along any axis (date, level, parameter, origin, ensemble, …)**

  - **Provision of multi-model data on a single grid (regridding to any lat/lon grid)**

  - **Sub-area selection**

  - **Reduces volumes to be downloaded by many order of magnitude**

ECMWF

TIGGE portal at ECMWF

TIGGE portal at ECMWF (using Ajax)

TIGGE Portal: based on WebMARS

TIGGE Portal: grid selection

TIGGE Portal: area selection

TIGGE Portal: based on WEBMARS

TIGGE Portal: use SMS for scheduling users' requests

- **Monitoring**
- **Per user limits**
- **Global limits**

TIGGE Portal: direct access to MARS

- **Most of the data is on tape: access to long timeseries**
- **Even finer control of resource usage**

```
marsth

name          Idle      Pid  Task    ID   Info    Request
------------------------------------------------------------------
cleaner          2     94546  547      0           -
flusher          4     68598  599      0           Waiting for connection
   hpss      10:53     68598  878      0           -
httpsvr          9     44498   36      0           Idle
mars             7     72778  778      0           Waiting for connection
   queue         7     72778  282      0           -
      queue   0:06     68118  118      0           -
   mars         30     72778  314      0           -
      mars    0:01     43156  692      0           -
   mars         15     72778  322      0           RETRIEVE 00000001_baudouin_raoult_ecmwf_int web03 TI ECMF CF SFC 20070601 0000
      mars       7     87224  224  3586081         Retrieving 19420 fields from 61 layouts 0%
   queue         7     72778  570      0           -
      queue      7     64812  348      0           -
   mars          7     72778  618      0           -
      mars    0:02     81798  798      0           -
   mars          6     72778  842      0           -
      mars       6    105786  788      0           -
marsadm       3:34     69510  510      0           Executing 'ps'
marsadm     28days     97532  532      0           Idle...
marsadm      2days     86642  642      0           Executing 'ps'
newops      10days     62814  350      0           Executing 'ps'
newops      10days     92438  438      0           Executing 'ps'
newops      40days     57080  616      0           Executing 'df'
newops      12days     61458  994      0           Executing 'ps'
reader           7     84546  546      0           Waiting for connection
   hpss         14     84546  146      0           Waiting for reader.t
      reader   13 H    74148  148  3586081  Read 0 Reading 70.1765 Mbytes from hpss 0%
         reader  13    74148  164      0     Wait 0 Double buffering 70.1765 Mbytes 0%
   hpss          0     84546  258      0           59 requests queued
   hpss         12     84546  930      0           Waiting for reader.t
      reader   12 H    78432  432  3586081  Read 0 Reading 70.1765 Mbytes from hpss 0%
         reader  12    78432  448      0     Wait 0 Double buffering 70.1765 Mbytes 0%
```

TIGGE Phase 2

- **Phase 2: distributed approach, higher development effort**
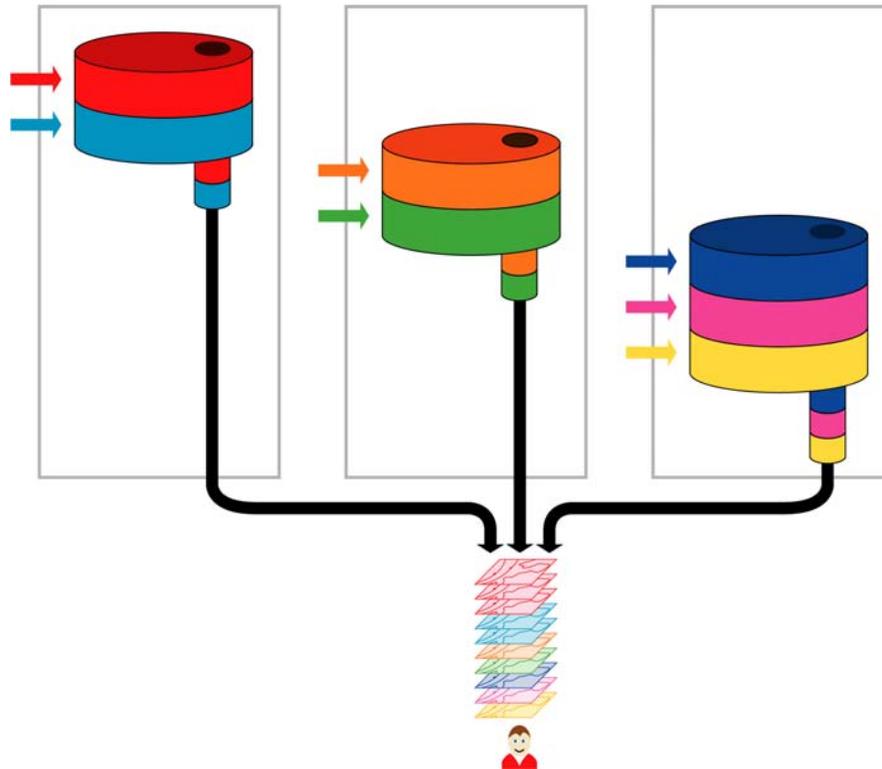
BRIDGE

- **Bridge is a 2 years project funded by the EC under the FP6-IST programme.**

- **It will demonstrate the benefits of GRID technology for international cooperation, in particular between Europe and China**

- **This work focuses on the development of interoperable Grid infrastructures (CNGrid, GRIA)**

- **Three applications: Aircraft Design, Meteorology, Drug Design**

ECMWF

BRIDGE meteorology application

- **BRIDGE project gives us the funding to explore how to implement TIGGE Phase 2**

  - Partners: ECMWF, DWD, CMA and CNIC (Computer Network Information Center, Chinese Academy of Sciences)

- **Creation of probabilistic weather forecasts products from the TIGGE dataset in a distributed fashion**

  - Each site hosts only part of the TIGGE data

  - Each site offers basic operations on the data (e.g. computing an average)

- **Strategy: minimize data transfers**

  - Run operations at data location

  - Decompose operations in simpler ones

  - Most of the time intermediate results are much smaller

ECMWF

Example: EPSGram

EPS Meteogram
Reading (48m) 51.46 °N 1.33 °W
Deterministic Forecast and EPS Distribution Thursday 8 March 2007 12 UTC
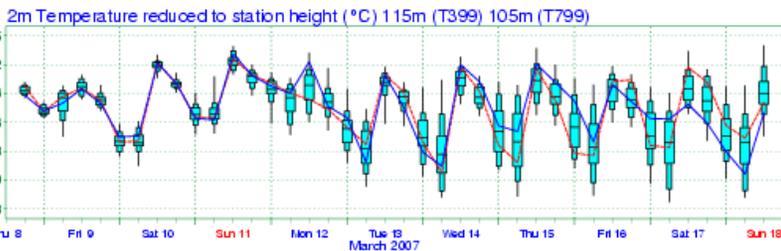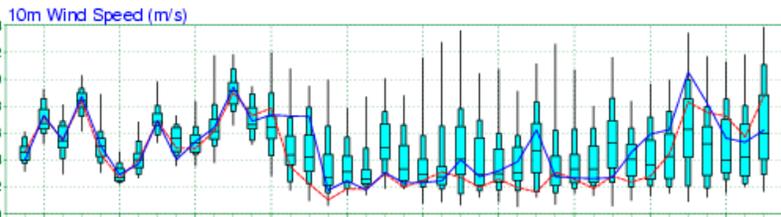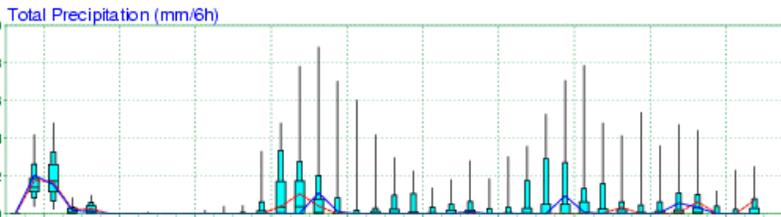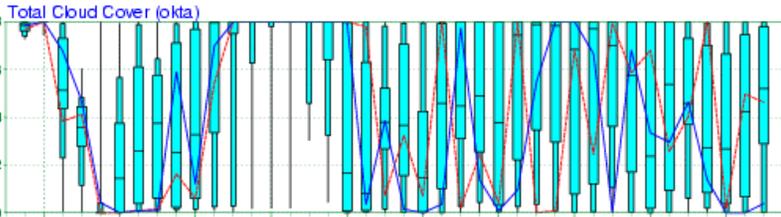
- **Send lat/lon location at each sites**

- **Receive list of values back**

- **Compute distributions**

- **Generate plot**
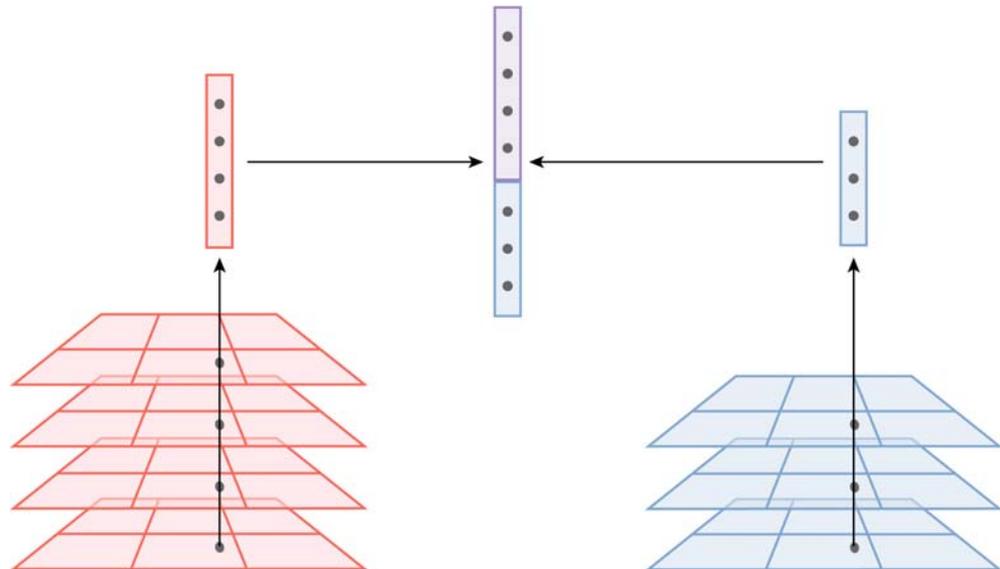
EPS Products

- **Examples**

  - **Ensemble mean**

  - **Standard deviation**

  - **Clustering**

  - **Probability of weather events**

  - **Extreme Forecast Index**

  - **EPSgram**

- **Some products can be decomposed in simpler operations on a subset of members**

  - **Ensemble means**

- **Some products need all the members**

  - **Clustering**

ECMWF

Example: Ensemble mean

- **Data requests describes 10 fields**

- **6 are available from ECMWF**

- **4 are available from CMA**

- **S1 = sum(6 ECMWF fields) performed at ECMWF**

- **S2 = sum(4 CMA fields) performed at CMA**

- **Intermediate results and associated metadata moved to site were user invoked request: (S1,6) and (S2,4)**

- **Final result A=(S1 + S2)/(6+4) computed locally and returned to user**

ECMWF

Decomposable operation

ECMWF/GRIA

CMA/GOS

Data repository

Data repository

Data service → Operation service

Data service → Operation service

Data

Result

Data

Result

Result 1   Result 2

Operation Service

Result

DWD/GRIA

ECMWF

Non-decomposable operation

ECMWF/GRIA

CMA/GOS

Data repository

Data service

Data

Data repository

Data service

Operation service

Data1

Data 2

Result
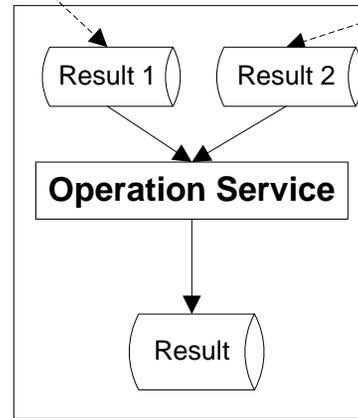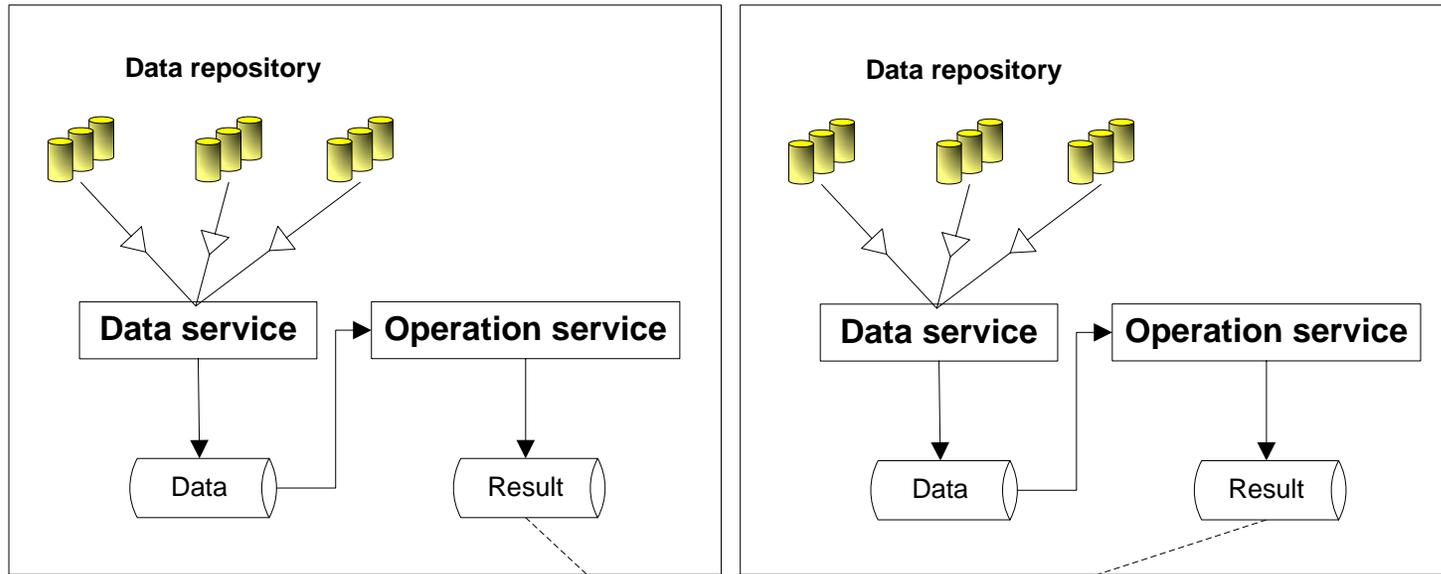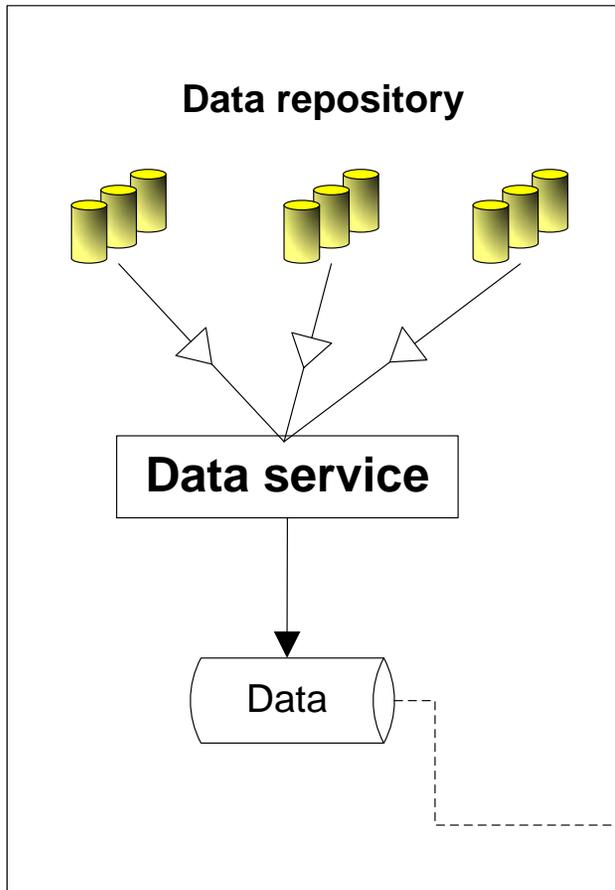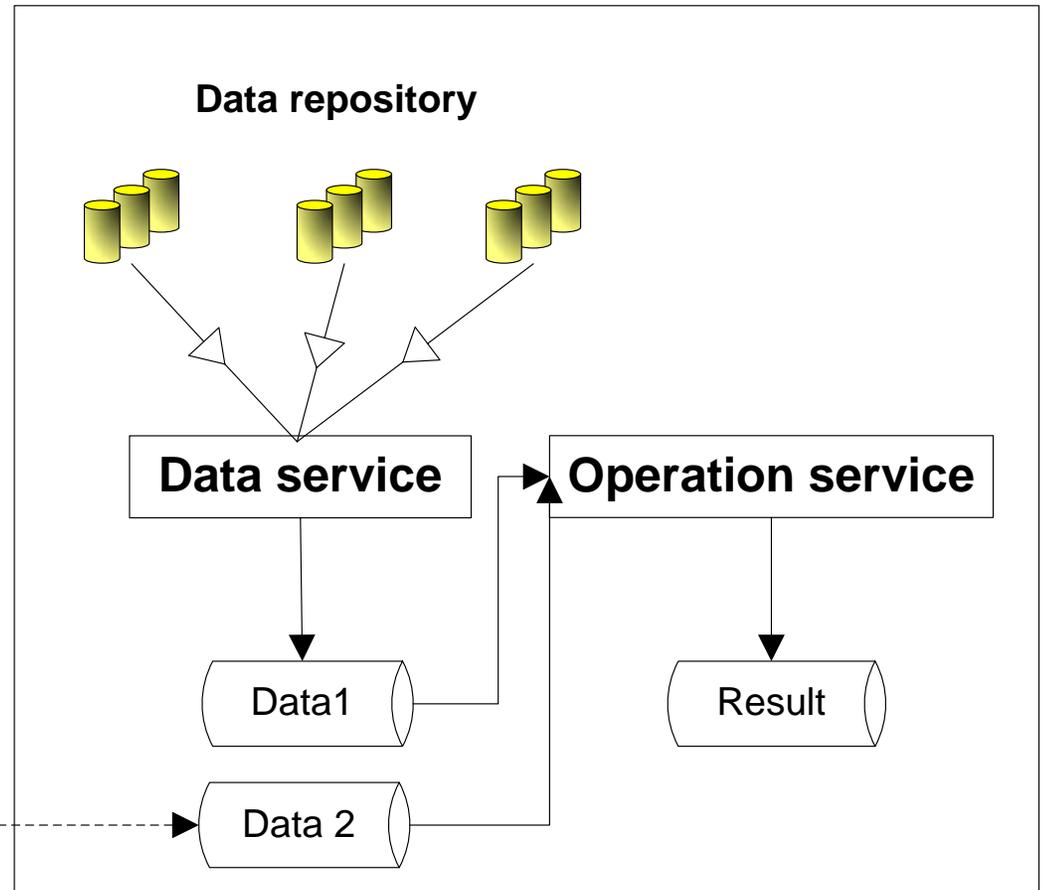
ECMWF

Products description

- **At ECMWF, most EPS products are computed using the MARS/Metview macro languages**

    - **Notion of fieldsets: A variable contains many fields**

    - **Computations are expressed as high level formulas:**

$$speed = sqrt(u*u+v*v)$$

- **In BRIDGE we will express the computations using a similar syntax**

    - **High level formulas…**

$$epsmean(date:20070101,param:total\_precipitation)$$

    - **… translate into low level formulas**

$$A = retrieve\_ecmwf(date:20070101,param:total\_precipitation)$$

$$B = retrieve\_cma(date:20070101,param:total\_precipitation)$$

$$E = (sum(A) + sum(B)) / (count(A) + count(B))$$

ECMWF

Example

plot(retrieve(origin:babj,…) + retrieve(origin:ecmf,…))
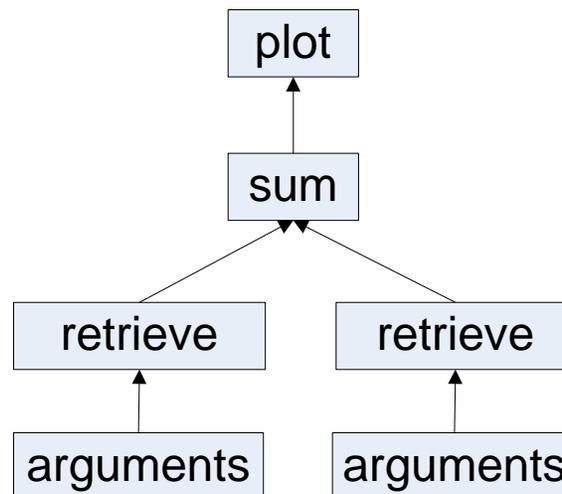
ECMWF
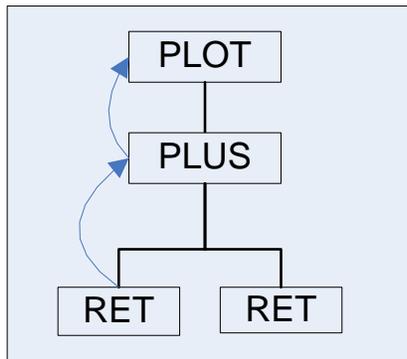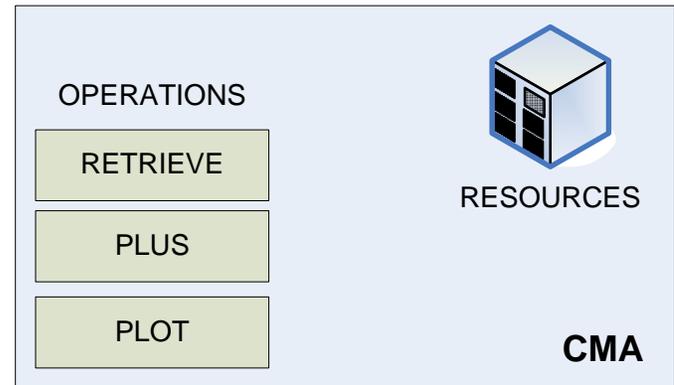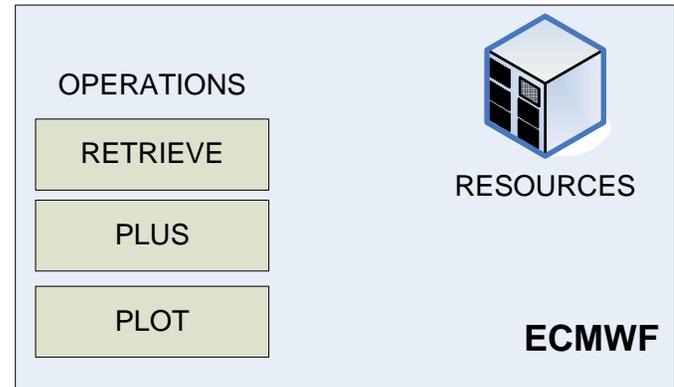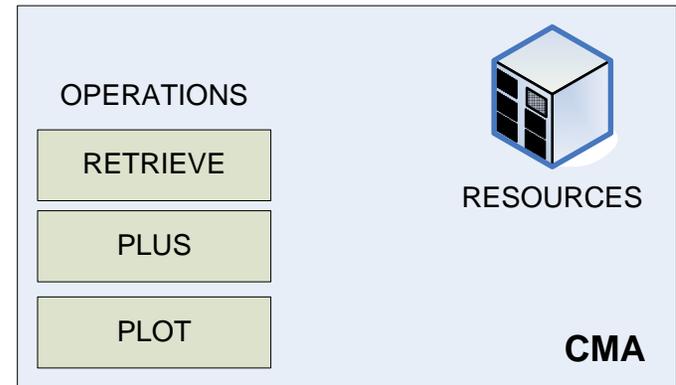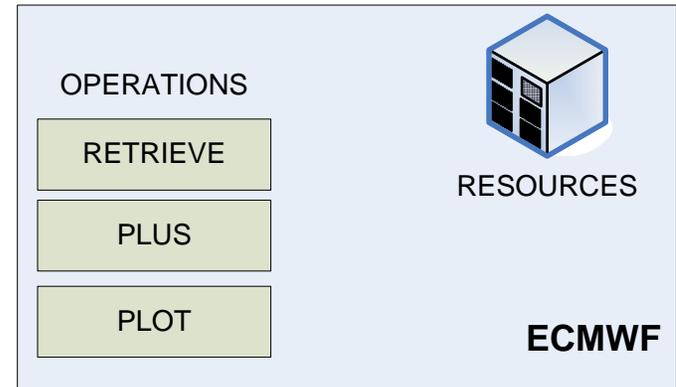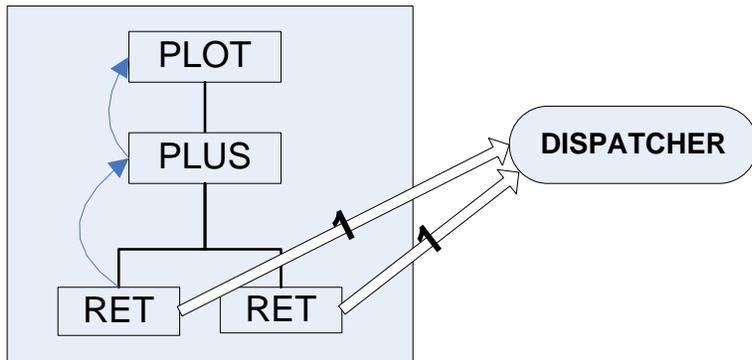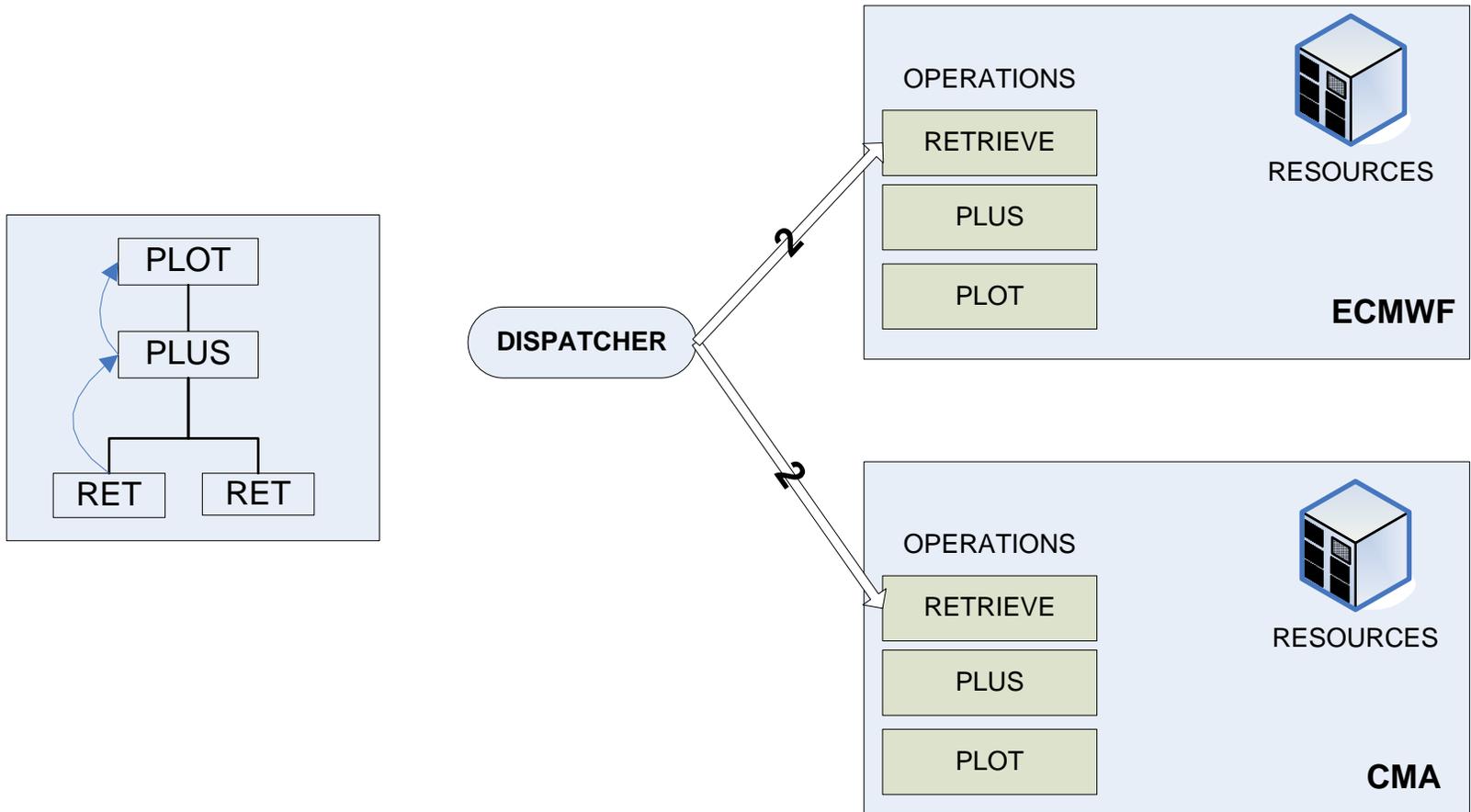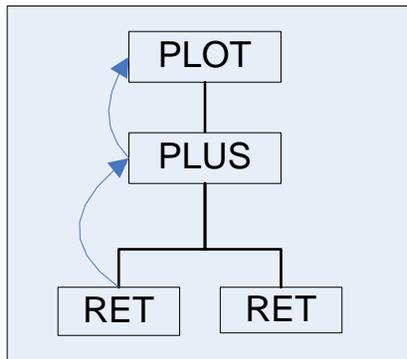
Example (cont.)

PLOT

PLUS

RET   RET

DISPATCHER

OPERATIONS

RETRIEVE

PLUS

PLOT

RESOURCES

ECMWF

OPERATIONS

RETRIEVE

PLUS

PLOT

RESOURCES

CMA

ECMWF

Example (cont.)

Example (cont.)

Example (cont.)

PLOT

PLUS

RET   RET

DISPATCHER

OPERATIONS
RETRIEVE
PLUS
PLOT

3
3
RESOURCES

ECMWF

OPERATIONS
RETRIEVE
PLUS
PLOT

3
3
RESOURCES

CMA
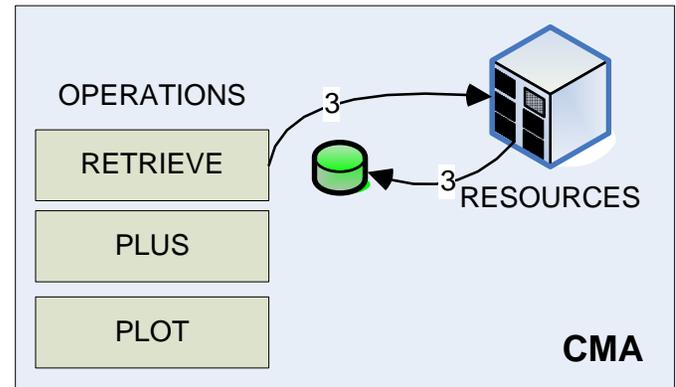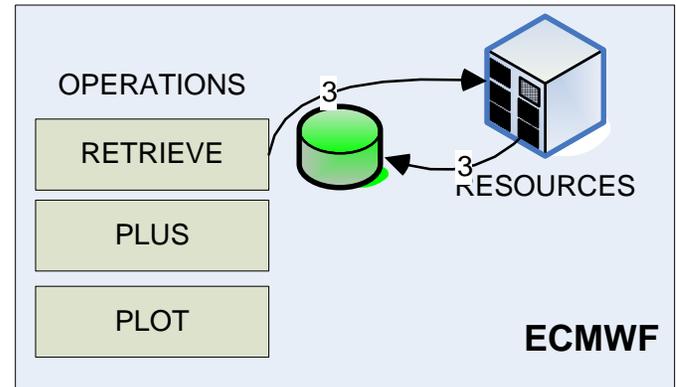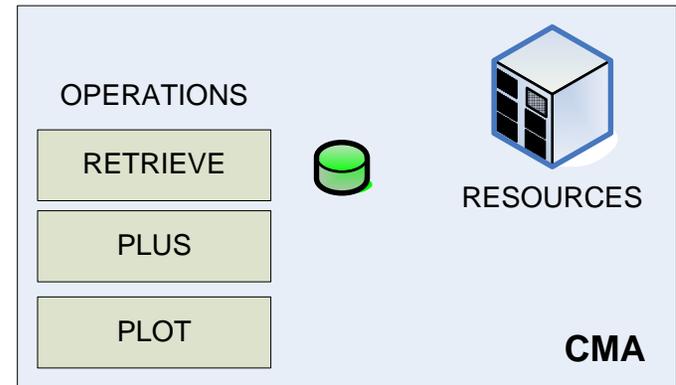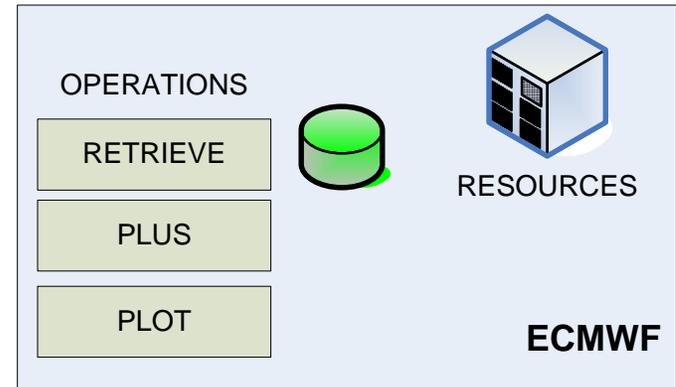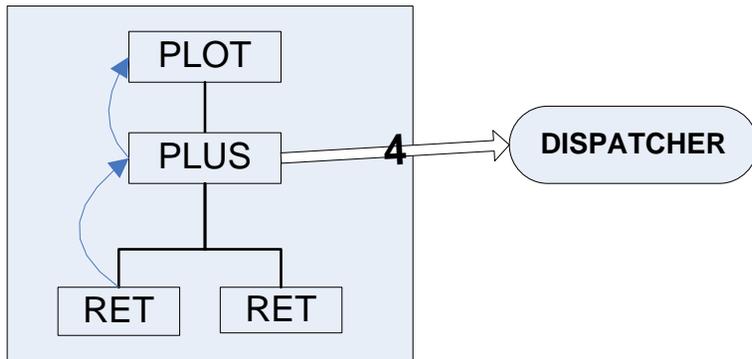
ECMWF

Example (cont.)
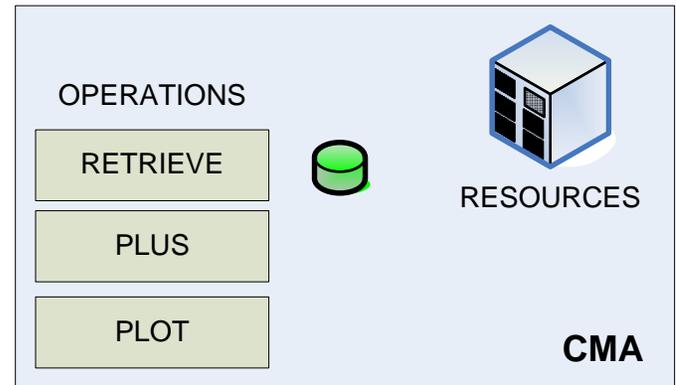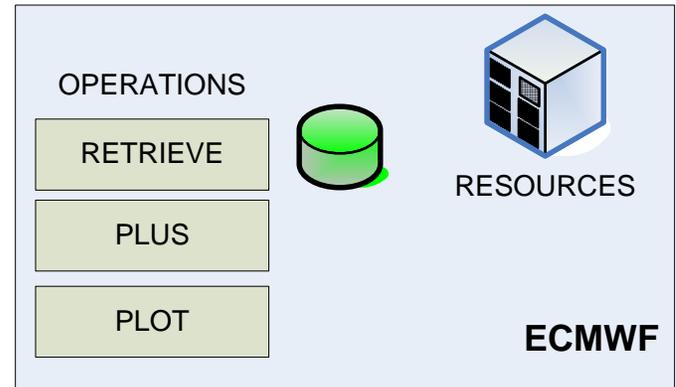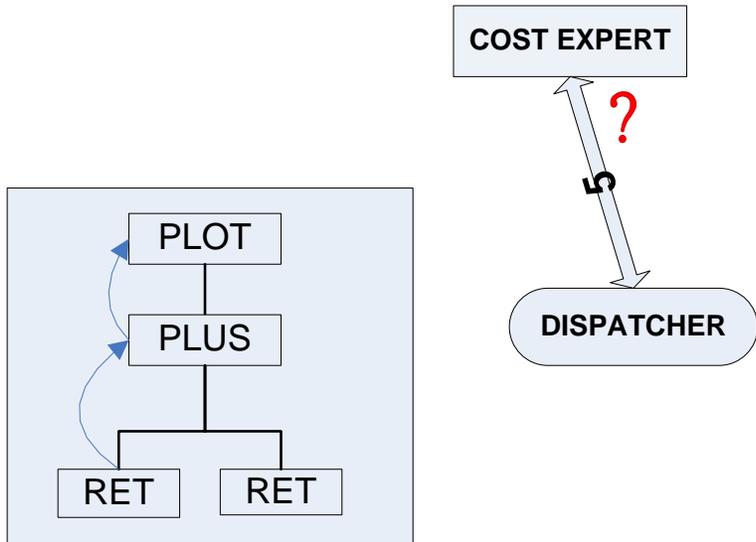
Example (cont.)
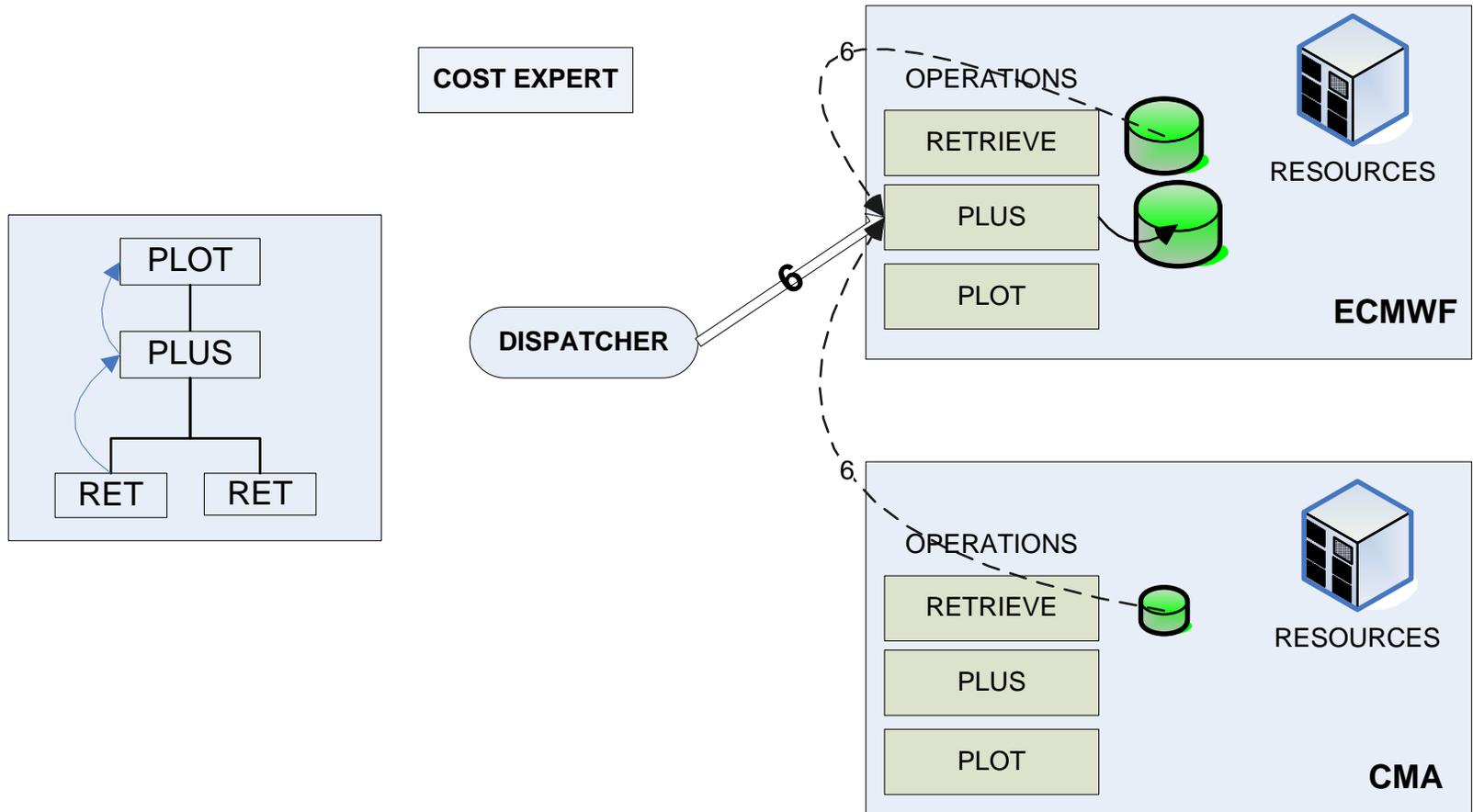
Example (cont.)
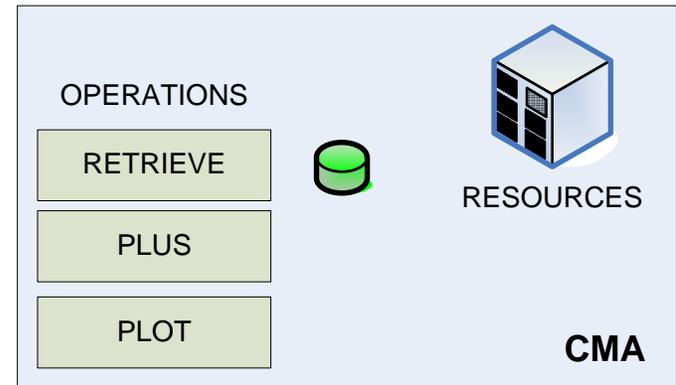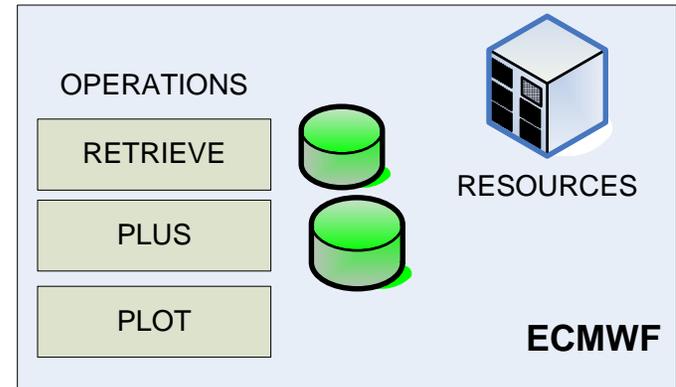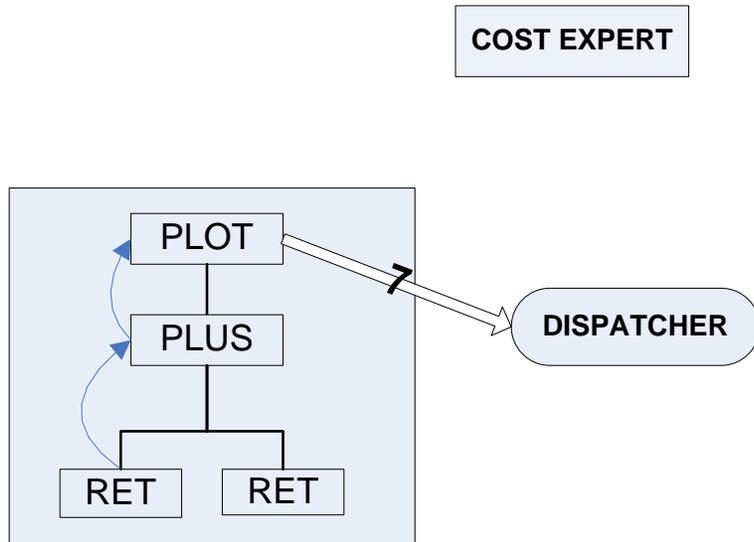
Example (cont.)
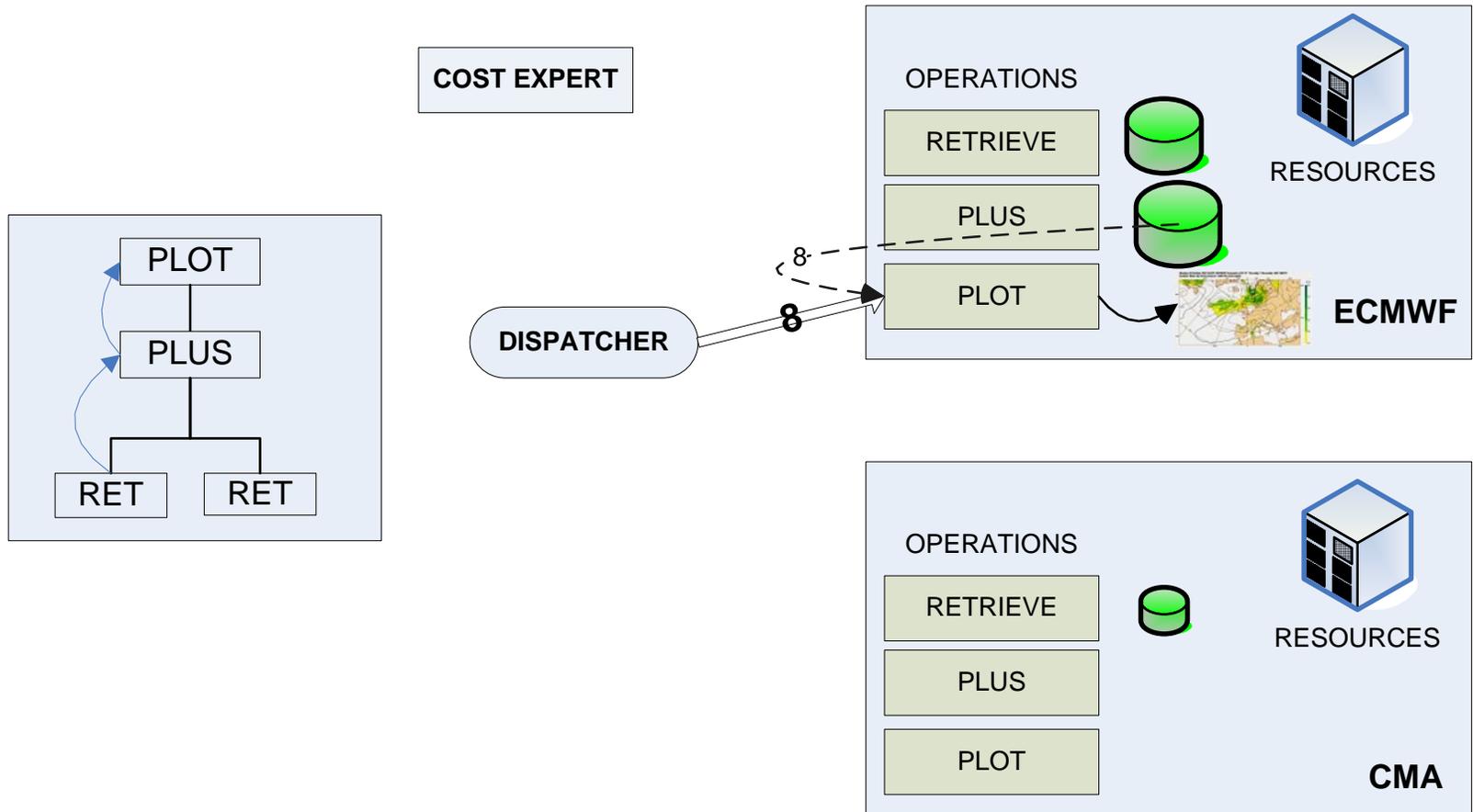
Example (cont.)

Example (cont.)
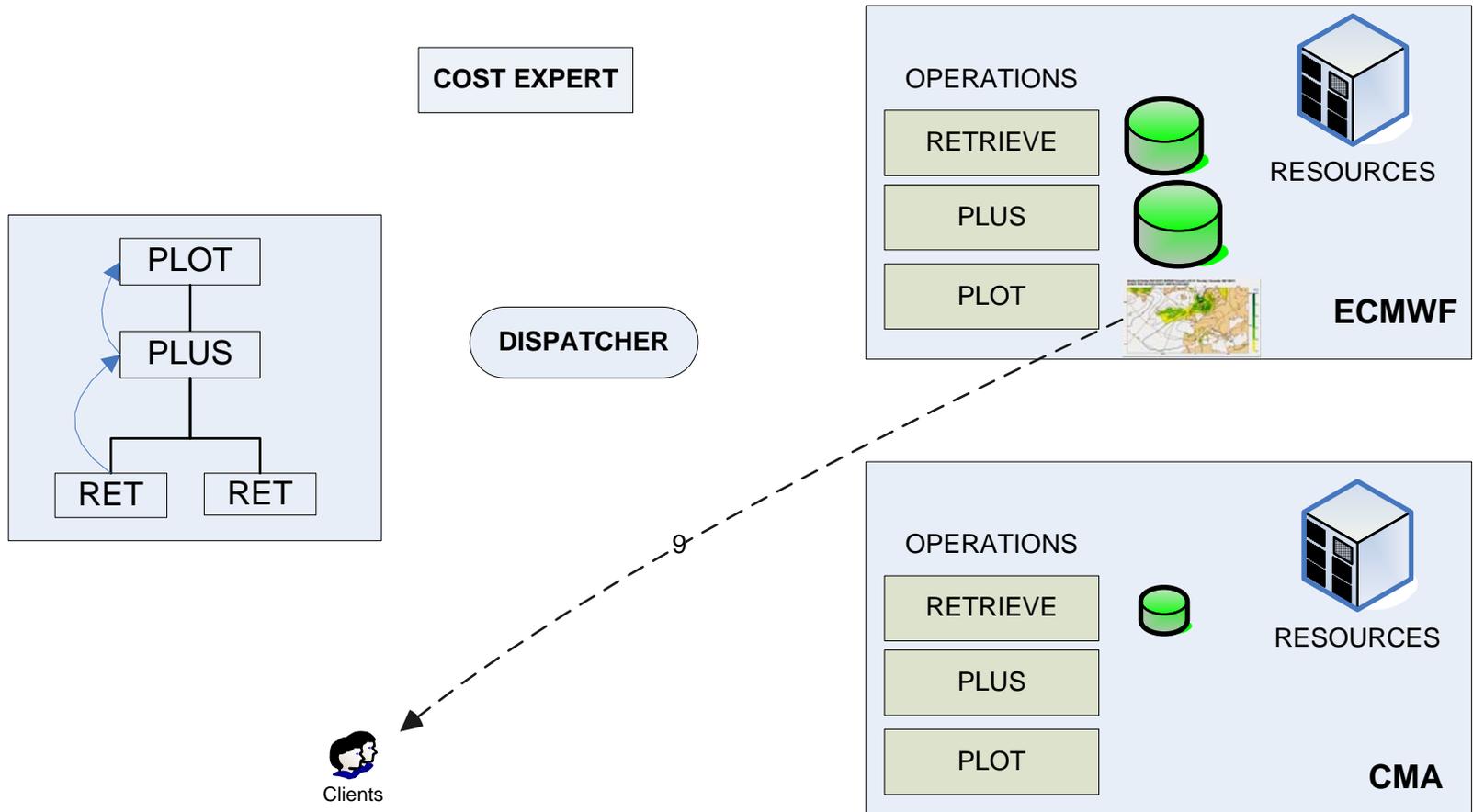
Conclusion

- **TIGGE Phase 1 is progressing well**
  - **Strong governance**
  - **Very good working relationship between CMA, NCAR and ECMWF**

- **BRIDGE project gives us the funding to explore how to implement TIGGE Phase 2**
  - **Study how to create EPS products in a distributed environment**

ECMWF