TIGGE– Infrastructure and Tools

TIGGE Workshop

ECMWF, Reading, UK

1 - 4 March 2005

Alfred Hofstadler

Meteorological Applications Section

ECMWF

TIGGE Infrastructure and Tools: Topics

- **Data**
- **Archive**
- **Access**
- **Tools**
- **Infrastructure**
- **Some questions**

TIGGE Infrastructure and Tools: Data

- **Data specification**

  - **Resolution**

  - **Levels, Level type**

  - **Parameters**

  - **Number of Ensembles**

  - **Initial times, Steps**

- **Data Format**

  - **GRIB2**

  - **BUFR (for non-gridded information; e.g. tracks)**

  - **Conversion APIs (GRIB1, NetCDF)**

  - **(WEB-)Graphics: SVG, PNG**

TIGGE Infrastructure and Tools: Archive

- **MARS**
  - GRIB/BUFR restriction
  - "Nice" extra features (e.g. interpolation, sub-area)
  - Large data amounts possible
- **File-based "Archive"**
  - Can store anything (e.g. project description documents)
  - Can be copied on tape drives, DVDs
  - WMO file naming convention
  - Space restrictions
- **Other solution**

TIGGE Infrastructure and Tools: Access

- **Centralised data repositories:**
  - Pull from ftp server via LAN/Internet (common procedures)
  - Data Server with (restricted) MARS access
  - Tapes, CDs, DVDs

- **Distributed: Data- and/or Application-GRID**

- **Access to full data sets or a subset (e.g. only last month)**

- **Security (e.g. registration, passwords)**

- **Data policies**

- **Common (graphical) interfaces and request procedures**

- **Time delay (near-realtime, days, weeks, months)**

- **Backup procedures**

ECMWF

TIGGE Infrastructure and Tools: Tools

- **Simple data access tools**
  - MARS tools (interpolation, sub-area,…)
  - Data conversion APIs
 - Post-processing APIs
- **Interactive applications**
  - Modify post-processing parameters (eg. target areas)
  - Run case specific applications (e.g. LAM EPS, sensitivity)
  - Set up user specific post-processing/verification
- **WEB site**
  - Centralized or distributed with common portal
  - What information and in what format
  - Who administrates

TIGGE Infrastructure and Tools: Infrastructure Example 1

ECMWF

Archive

Processing

CMA ?

Archive
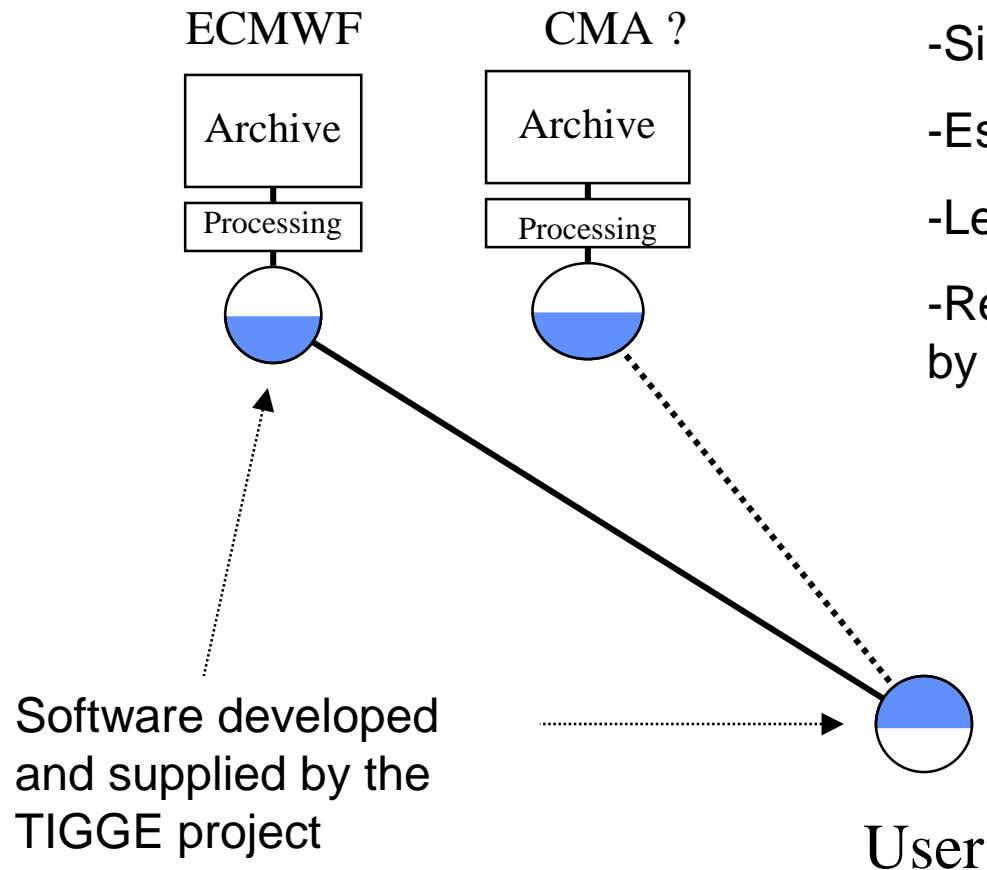
Processing

-Simple design

-Essentially a remote MARS access

-Less software development

-Retrieval and processing capacity provided by archive sites (e.g. ECMWF,CMA)

Software developed and supplied by the TIGGE project

User

TIGGE Infrastructure and Tools: GRID
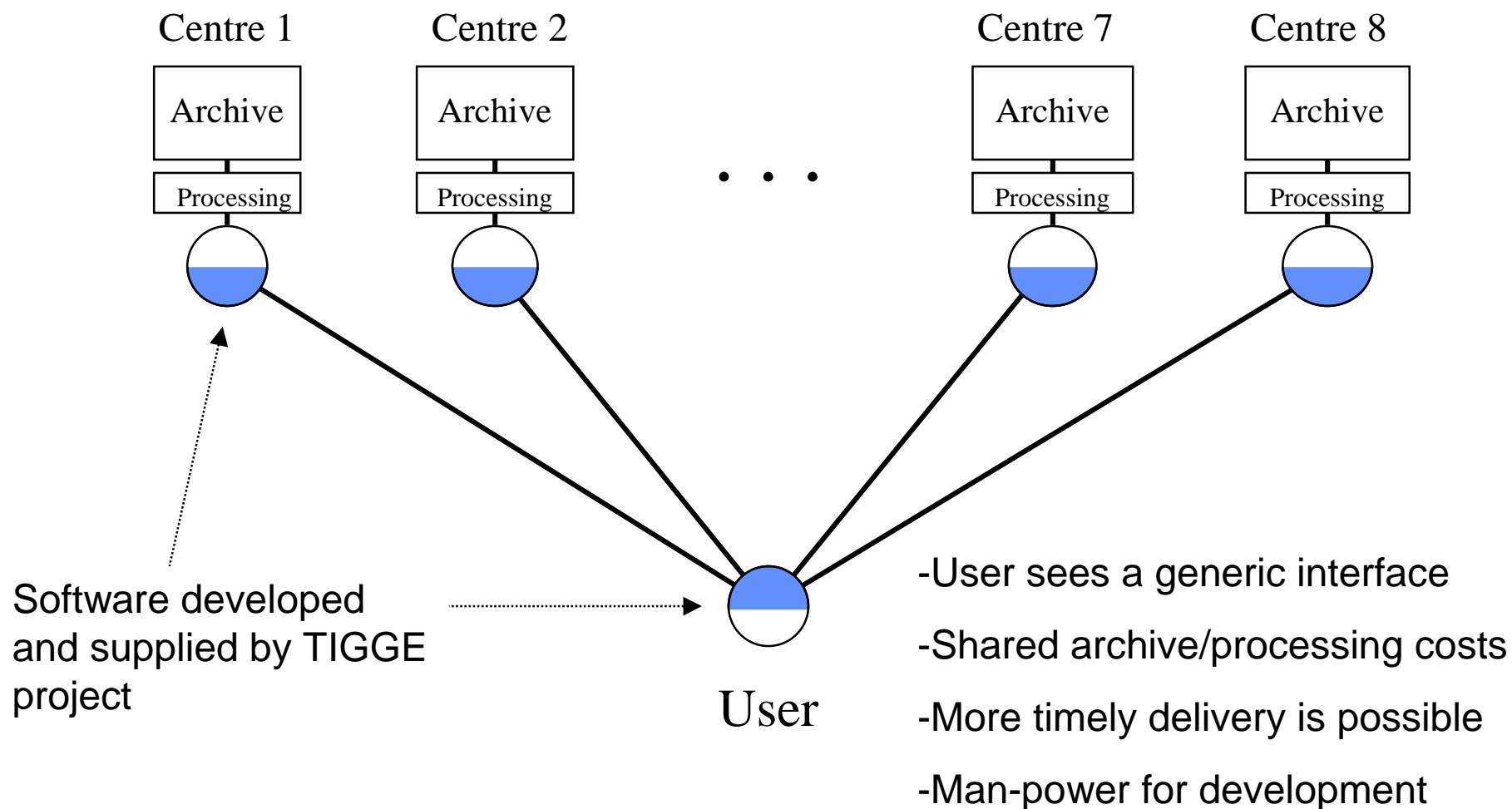
- **Data GRID**

  - **Distributed DBs with homogeneous catalogues**

  - **Data caching, duplication**

  - **WMO: FWIS (SIMDAT) standards/infrastructure**

- **Application GRID**

  - **Postprocessing**

  - **Extra model runs**

  - **Control EPS runs**

  - **Targeted EPS LAMs**

ECMWF

TIGGE Infrastructure and Tools: Infrastructure Example 2

Centre 1   Centre 2   Centre 7   Centre 8

| Archive | Archive | . . . | Archive | Archive |

Processing   Processing   Processing   Processing

Software developed
and supplied by TIGGE
project

User

-User sees a generic interface

-Shared archive/processing costs

-More timely delivery is possible

-Man-power for development

TIGGE Infrastructure and Tools: Example

- ## ECMWF EPS at T255

  - Selected pressure levels & surface fields   25 GB/day

  - Control forecast and initial conditions   8 GB/day

  - T255 to T399 implies a factor of 2.5 for field sizes

  - Figures are for two runs per day out to 10 days

  - Archive volumes: 82.5 GB/day

- ## TIGGE with up to 10 EPS producing centres

  - Assume higher resolution and more members at ECMWF

  - 5 times the ECMWF volumes may cover the needs

  - ECMWF's EPS is already archived; use a factor of 4 for TIGGE

  - 342 GB per day, 125 TB per year

- ## Does NOT include data required for regional models (boundary data on model levels) or wave data

TIGGE Infrastructure and Tools: Example

- **Receive 342 GB per day via the Internet**

  - 32 Mbps sustained throughout the day

  - Will require a dedicated 100 Mbps link to allow for variability

  - ECMWF has currently a 60 Mbps Internet link

  - Shipping of tapes would be a cumbersome procedure with long delays

- **Archive 125 TB per annum**

  - There is probably no need to keep backups of the TIGGE data (the originating Centres or a mirror site should do so)

  - The projected ECMWF archive growth for 2005 is ~ 1.4 PB (inclusive backup copies)

  - TIGGE would be ~10% of the primary data growth

TIGGE Infrastructure and Tools: Some questions

- **Which sites wish to provide a TIGGE service (NCEP, BMRC,…)**

- **What data should be exchanged and archived:**
  - resolution, levels, level type, parameter, number of ensembles, initial times, steps

- **What Data Formats will have to be supported:**
  - GRIB2, BUFR, GRIB1, NetCDF

- **How to provide access to data: centralized data repositories, distributed GRID, standard interfaces, WEB site**

- **What will be the typical request: complete fields, sub-areas**

- **What time-delay for data provision is acceptable**

- **What tools and applications are needed**

- **Funding for hardware, media, software development**