THORPEX Interactive Grand Global Ensemble (TIGGE)

Baudouin Raoult

ECMWF

Key TIGGE objectives

- **To enhance international collaboration between operational centres and universities on development of ensemble prediction**
- **To develop new methods to combine ensembles from different sources and to correct for systematic errors (biases, spread)**
- **To increase understanding of the contribution of observation, initial and model uncertainties to forecast error**

Key TIGGE objectives

- **To evaluate the feasibility of an operational interactive ensemble system that responds dynamically to changing uncertainty and that exploits new technology for grid computing and high-speed data transfer**
- **To evaluate the elements required of a TIGGE Prediction Centre to produce ensemble-based predictions of high-impact weather, wherever it occurs, on all predictable time ranges**
- **To develop a prototype future Global Interactive Forecasting System**

TIGGE Participation

- **Already several operational global EPS**
  - **BMRC**
  - **CMA**
  - **CPTEC**
  - **ECMWF**
  - **FNMOC**
  - **JMA**
  - **KMA**
  - **MSC**
  - **NCEP**
  - **Expressions of interest from these and others (Met Office, …) to contribute to TIGGE**
- **Other groups already interested in using TIGGE data**

TIGGE: Work So Far ...

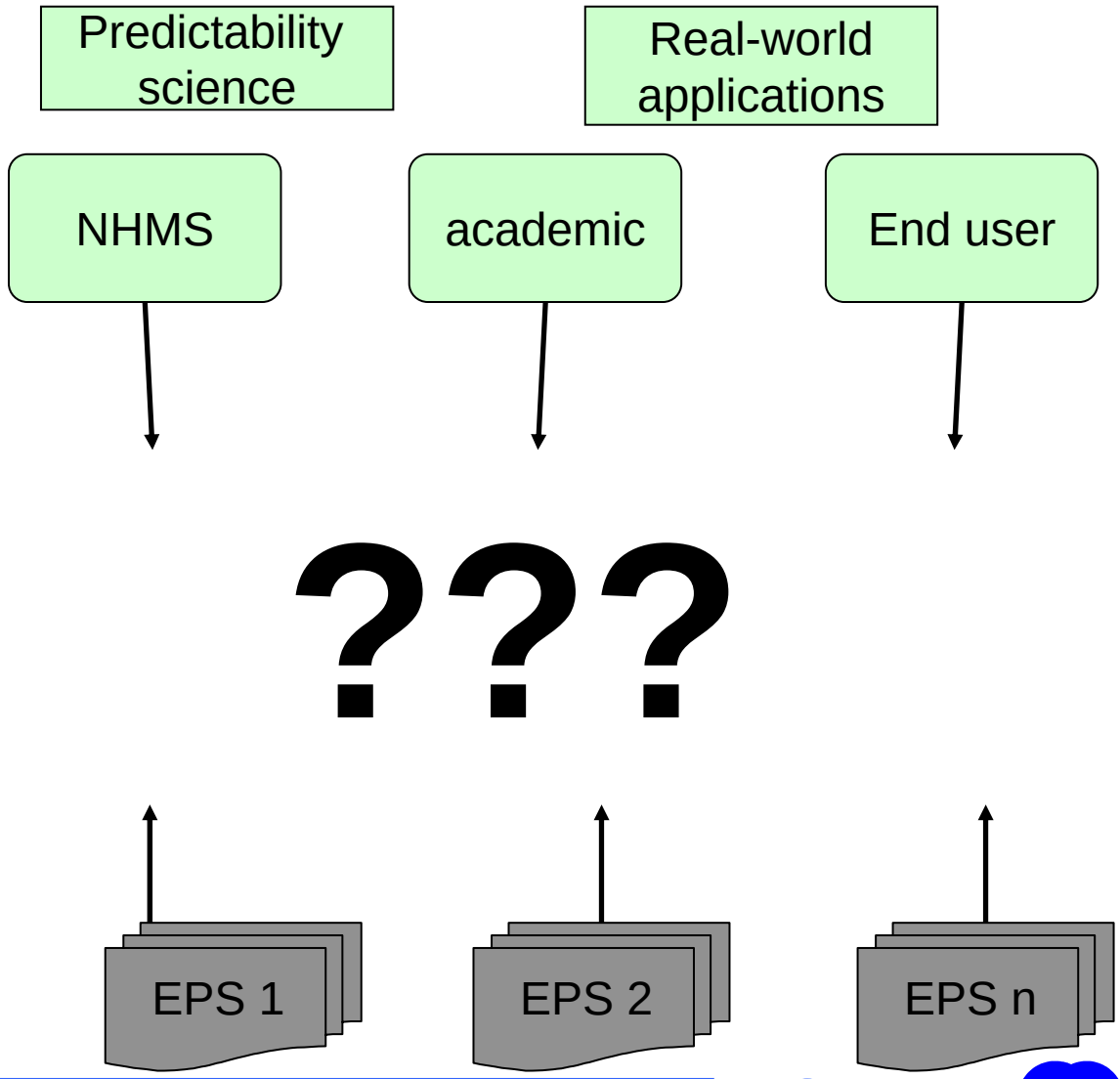
- **First TIGGE Workshop**
  - ➔ **ECMWF 1-3 March**
  - ➔ **Requirements**
- **Working group on archiving**
  - ➔ **ECMWF 19-21September**
  - ➔ **Technical implementation by Archive Centers**
- **Working group on implementation**
  - ➔ **ECMWF 10-11 November**
  - ➔ **Feedback from Data Providers**
- **GIFS-TIGGE Working Group**
  - ➔ **NCAR, 15-16 November**
  - ➔ **Scientific aspects, Data policies, …**

1st TIGGE Workshop (1-3 Mar.)

- **Address strategy to achieve TIGGE objectives**
- **Focus on user-requirements and infrastructure needed to meet these**
- **Produce outline plan and timetable**
- **60 participants from operational centres and universities worldwide**
- **Report submitted to THORPEX Executive Board and International Core Steering Committee**
- **GIFS/TIGGE Working Group appointed to oversee TIGGE**

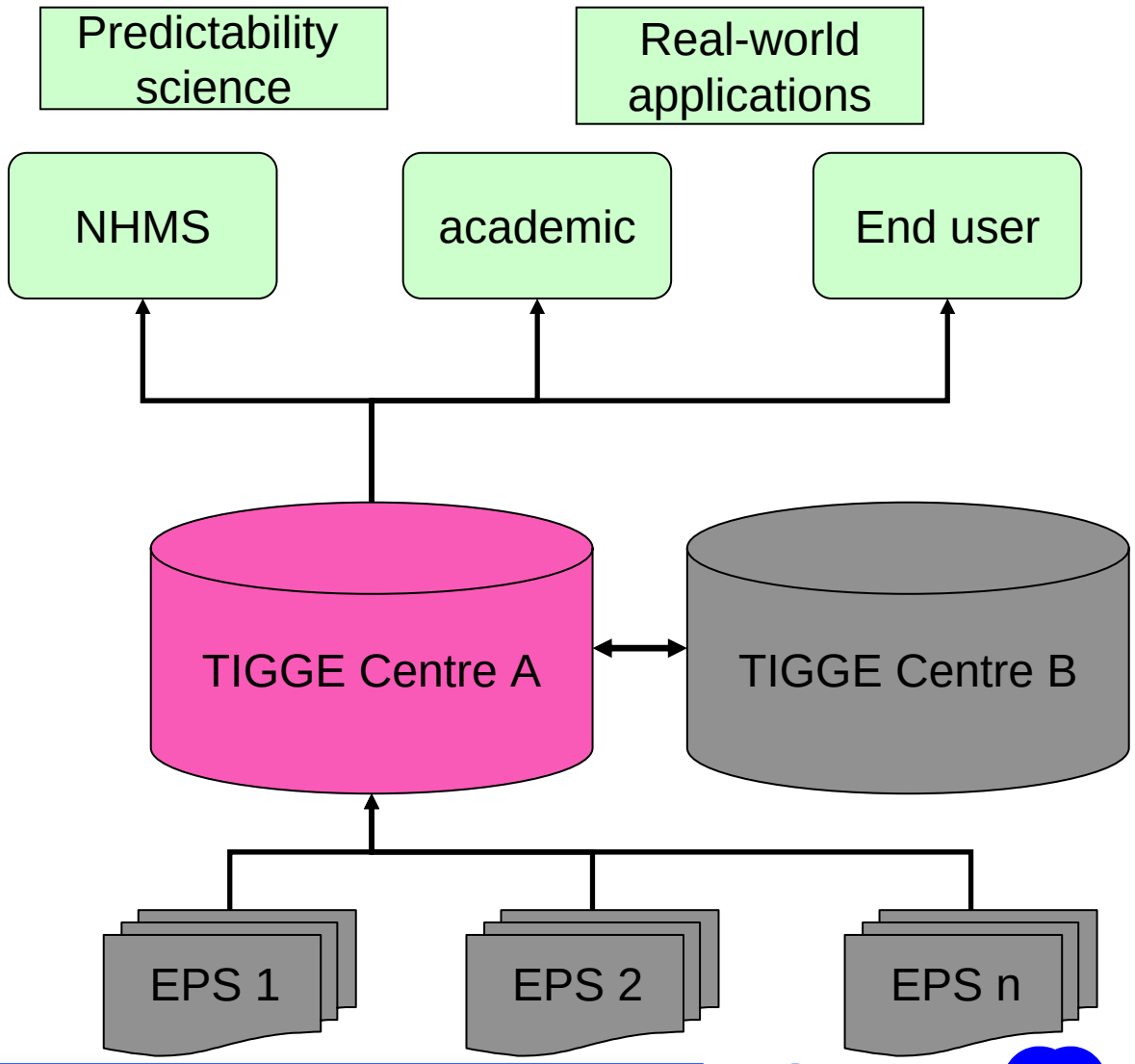
1st TIGGE Workshop: Aims

- **Users**
  - **Who?**
  - **What for?**
  - **How?**
  - **When?**
- **Requirements**
  - **What data?**
  - **Format?**
  - **How to access?**
- **Contributors**



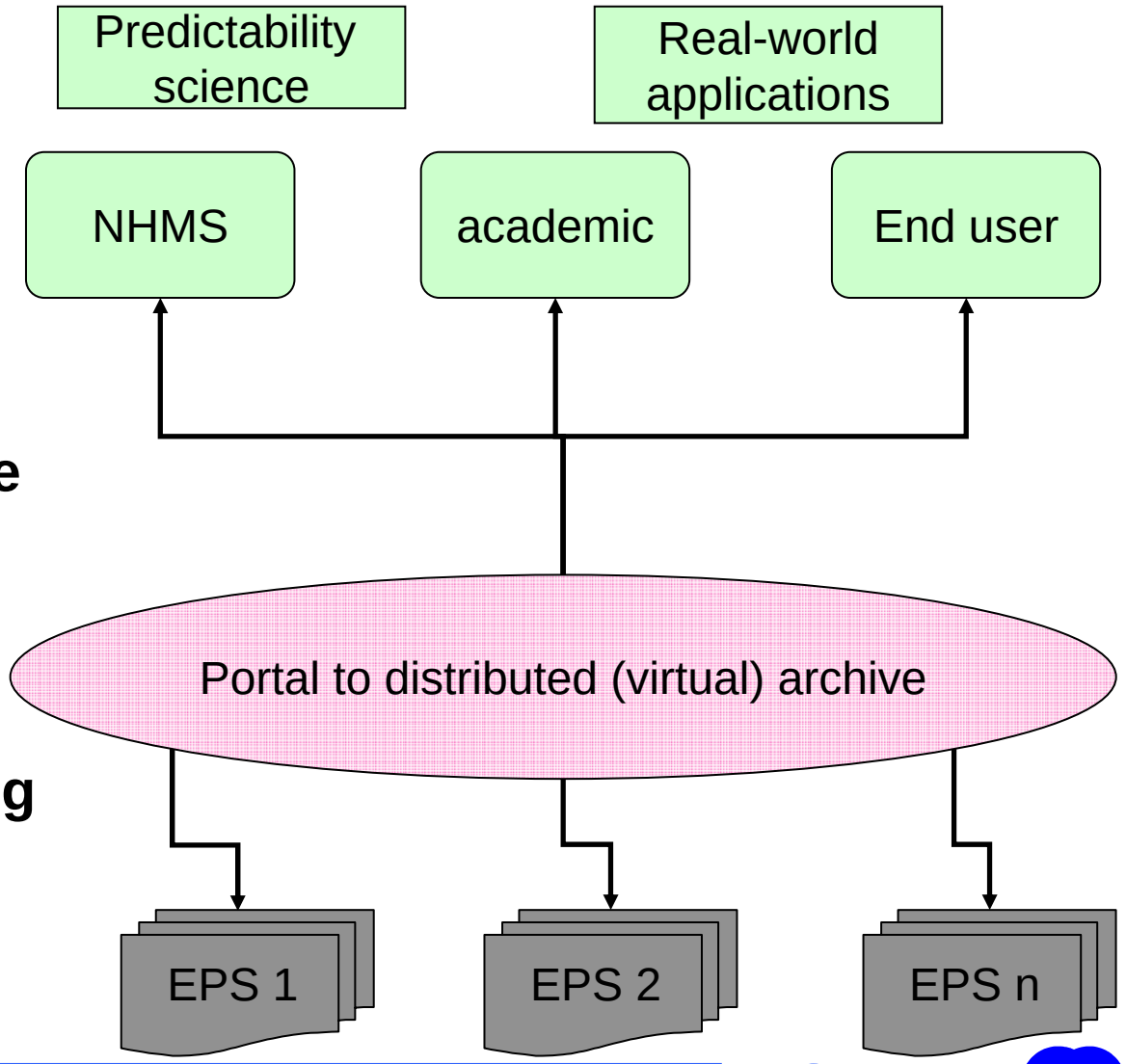
TIGGE infrastructure Phase 1

- **Data collected in near-real time (via Internet) at central TIGGE data archives**
- **Can be implemented now at little cost**
- **Can handle current data volumes within available network and storage capabilities**



TIGGE infrastructure Phase 2

- **Data distributed over several repositories**
- **But keep efficient and transparent user access**
- **Flexible – minimise data transfers**
- **Needs substantial software development**
- **Coordination with WMO Information System**
- **Requires additional funding**

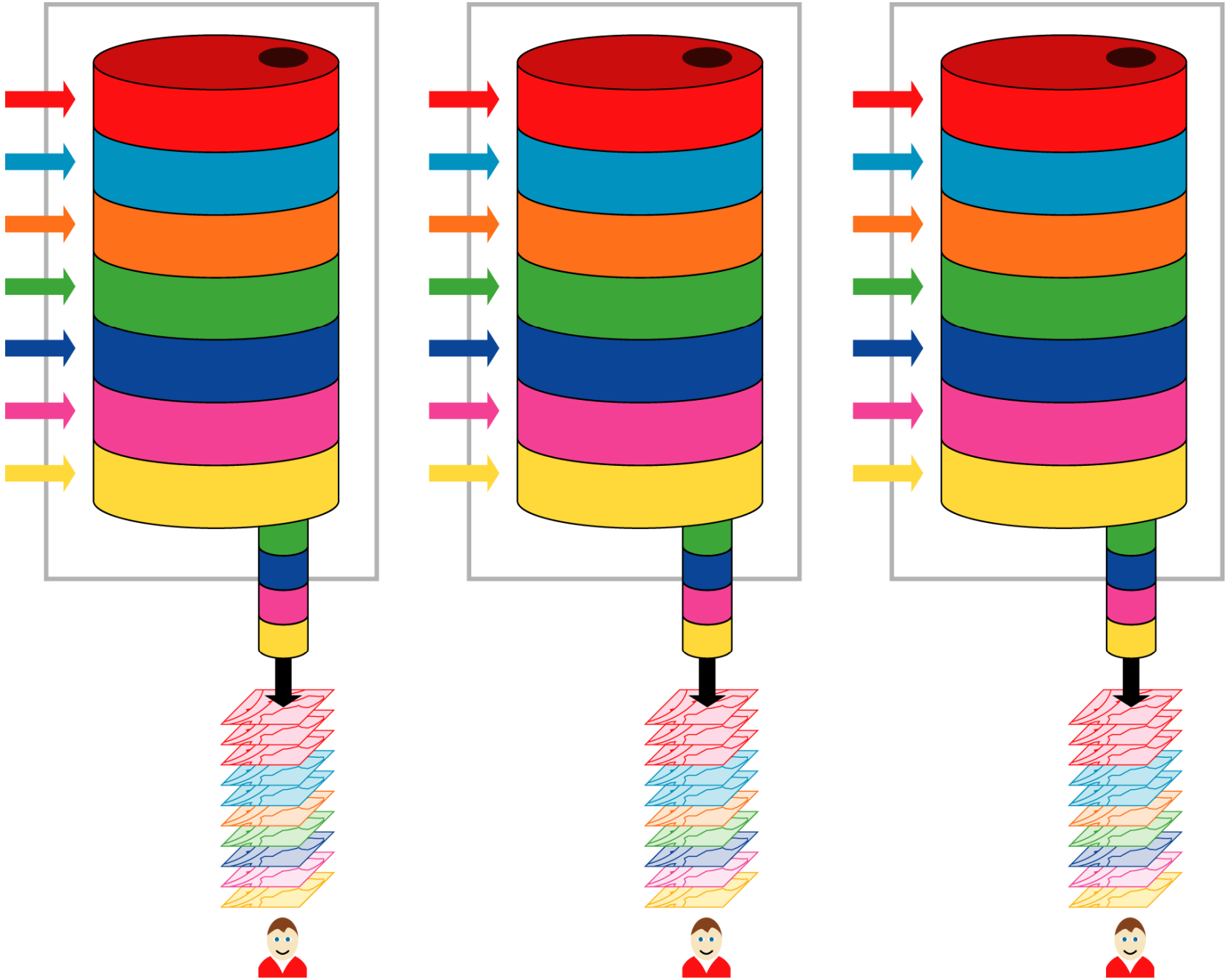


TIGGE Archive Working group (19-21 Sep.)

- **Aim**
  - **Phase 1 implementation**
  - **Reviewing the requirements set by the first TIGGE Workshop**
  - **Considered the various technical issues and solutions.**
- **Focused on building the database:**
  - **Interaction between Data Providers and Archive Centres**
- **Participants**
  - **CMA**
  - **CMC (for NAEFS)**
  - **ECMWF**
  - **NCAR**
  - **NCEP (Zoltan Toth on the phone)**

TIGGE Phase 1 – A summary

- **3 Archive Centres**
  - ➔ **CMA, NCAR, ECMWF**
- **8 Data providers (?)**
  - ➔ **NCEP, ECMWF, UKMO, JMA, BMRC, CPTEC, KMA, MSC**
- **Each Archive Centre will receive data from all the Data Providers**
  - ➔ **In near realtime**
- **Users will be able to get the same data from any of the Archive Centres**
- **No extra resources!**
  - ➔ **Use existing infrastructure**



Homogeneity of the TIGGE database

- **Homogeneity is paramount for the succeed of TIGGE.**
  - ➔ **The more consistent the archive the easier it will be to develop applications.**
- **A successful example: the DEMETER project**
  - ➔ **A multi-model seasonal forecast project**
  - ➔ **The effort put into creating a homogenous archive led to a variety of useful applications.**

There are three facets to homogeneity:

- **Common terminology**
  - ➔ **All fields should be described with the same attributes (dates, level, step, parameter, etc.)**
- **Common data format**
  - ➔ **All partners must agree on a common data format**
  - ➔ **They should also agree to use the same units**
- **Definition of a core dataset**
  - ➔ **When using fields to create a “grand ensemble”, i.e. when considering all members from several origin centres as a super ensemble, there must be an overlap (levels, time steps, parameters, …)**
  - ➔ **All Data Providers must adhere to the core dataset definition**

Data format: File structure

- **Use WMO file naming convention**
- **The partners agreed on using a common file structure**
  - ➔ **Two files (single level and pressure level) per time-step, all members, all parameters, all levels, in a given order**
  - ➔ **File sizes are intended to be optimum for network transfer, large enough for efficient mass storage operations**
- **Non-compliant files will not be accepted by the Archive Centres.**
- **Will add development work for Data Providers, but:**
  - ➔ **Will make data manageable at the Archive Centres**
  - ➔ **Will be beneficial to the users.**

Data format: GRIB edition 2 (GRIB2)

- **Only WMO standard that supports EPS data**
  - ➔ **The NAEFS community is committed to using it.**
- **There is little experience in using GRIB2**
  - ➔ **Which also means that there is no proliferation of local tables or extensions yet.**
- **The Archive Centres will:**
  - ➔ **Provide clear guidelines (best practices) on how all TIGGE fields should be coded in GRIB2.**
  - ➔ **Identify the list of GRIB2 codes, tables and templates to use for each of the fields of the TIGGE database.**
- **Existing application will have to be adapted to handle GRIB2.**
  - ➔ **GrADS, NCL, IDL, Ferret, Metview, etc…**

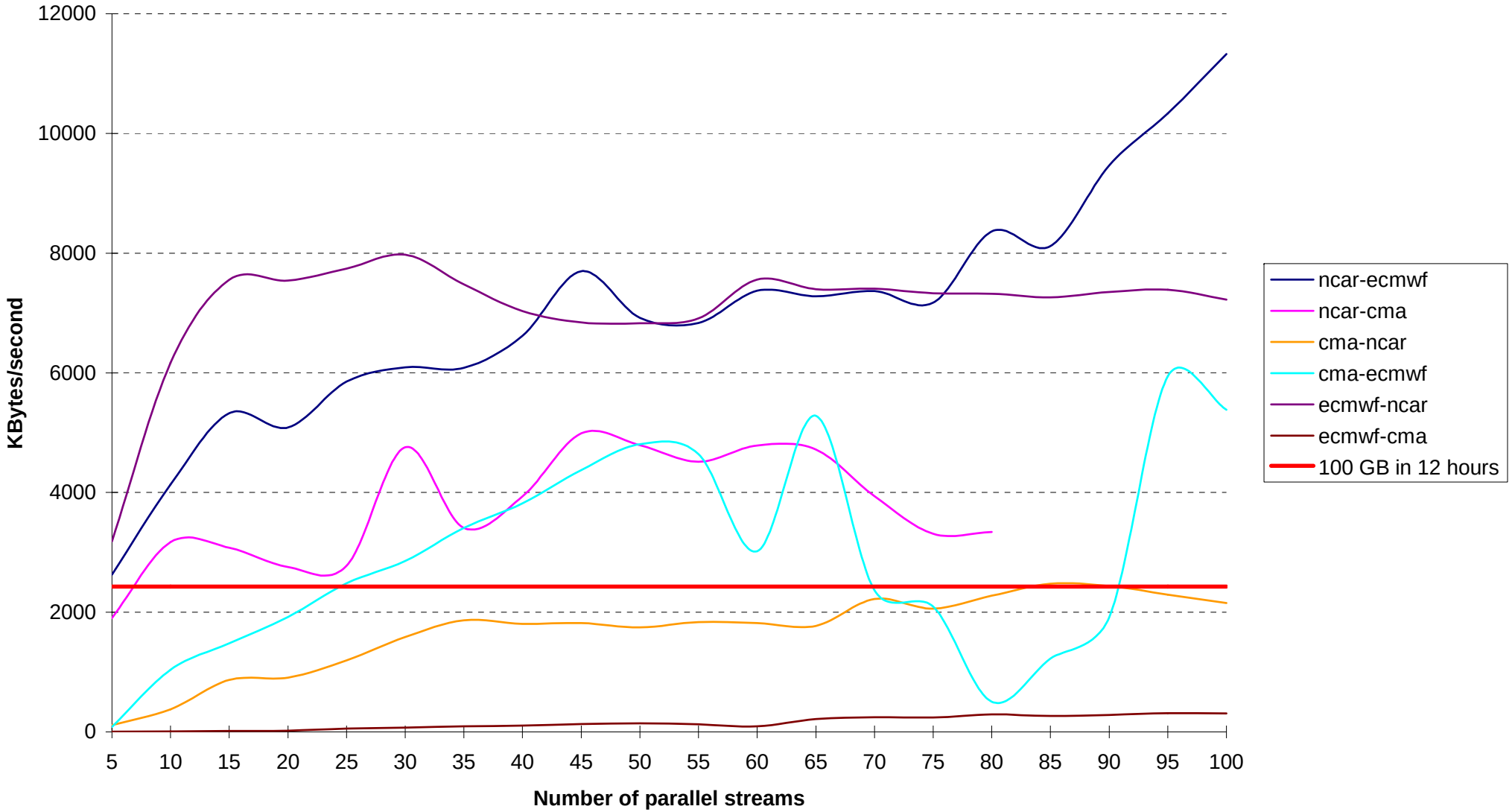
Data format: Grid and resolution

- **Preserve native grids and resolutions**
- **Data Providers to supply interpolation routines for conversion to regular lat-lon grids and for point extraction**
- **Archive Centres to specify interfaces for interpolation routines**
- **Archive Centres may endeavour to return data in regular grids using these interpolation routines**

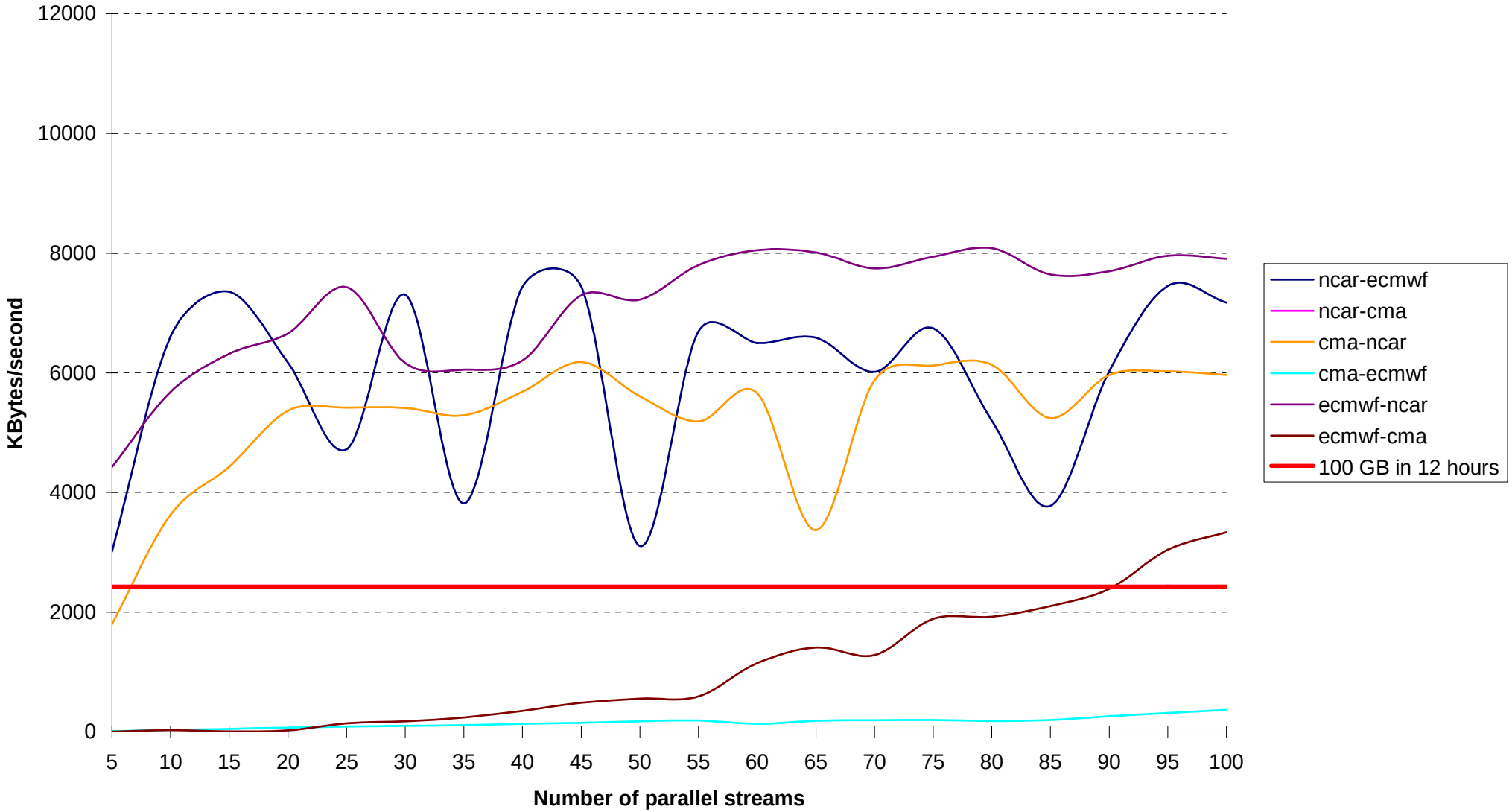
Network Bandwidth

- **Is a major risk**
- **Bilateral connections need testing**
  - **Between Archive Centres**
  - **Data Provider to Archive Centres**
- **Requires a lot of tuning/tweaking**
  - **Get vs. put**
  - **Number of parallel streams**
  - **Buffer sizes**
  - **TCP window sizes**
- **Target aggregate bandwidth: 100GB per 12 hours**
  - **Gives a chance to resend everything on the same day**

Aggregate rate using FTP *get* (client - server)



Aggregate rate using FTP *put* (client - sever)



Data flow

- **Data flow refers to the transfer of model output in near real-time from Data Providers to Archive Centres (not with the users)**
  - **Near real-time: in less than 24 hours**
- **The working group assessed several scenarios**
  - **FTP: Many to One**
  - **FTP: Many to Many**
  - **Specialised software: Unidata's LDM or DWD's AFD**
- **Preferred solution: LDM**
  - **In does not fulfil requirements, use Many to Many transfers**

Organisation of the collaboration

- **The success of TIGGE is directly linked**
  - ➔ **The degree of commitment of the partners**
  - ➔ **The ability of the partners to work together.**
- **Archive Centres will take the technical coordination**
  - ➔ **They will have a global view of the data production**
- **Each partner must nominate two contact points**
  - ➔ **A technical contact person**
  - ➔ **A scientific contact person**
- **Tools: email lists, web site**
  - ➔ **ECMWF offers to host a web site and mailing lists**

Quality assurance: Data integrity and completeness

- **Data corruption will be unavoidable**
  - **Large amounts of data will be moved across different media (memory, disks, network, tapes)**
  - **Use of checksums**
- **Issue of completeness**
  - **The objective is to have 100% complete data at the Archive Centres.**
  - **In real world this may not be achieved**
  - **Unfortunately, an incomplete dataset is often difficult to use.**
  - **Most of the current tools used for ensemble data assume a fixed number of members from day to day. These tools will have to be rewritten**
  - **Data Producers to endeavour resending missing data when possible**

Data retrieval

- **No unified access in Phase 1**
  - ➔ **ECMWF will utilise the MARS system**
  - ➔ **NCAR will build upon its Research Data Archive and Community Data Portal.**
  - ➔ **CMA is still in the development process of their data delivery system.**
- **Over time and with additional project support, it is expected that there will be opportunities to further unify the user interface by leveraging developments from the WMO Information System (WIS) effort.**

Risks

- **Use of the Internet (Low)**
  - ➔ **Building an operational system on the Internet may be difficult, as we have little control over it.**
- **File structure (Low)**
  - ➔ **Imposing a file structure may put too much burden on Data Providers and may discourage them.**
- **Using IDD/LDM (Medium)**
  - ➔ **This dissemination system may not work with firewalls, or may not be suitable for operations (traceability of problems, timeliness, …)**
- **Low bandwidth (Medium)**
- **GRIB2 (High)**
  - ➔ **There are currently very few tools able to handle GRIB2, in particular EPS data.**

Working group on Implementation (10-11 Nov.)

- **To address technical issues raised by the two preceding meetings.**
- **Participants from archive centres and data providers**
  - **BMRC**
  - **CMA**
  - **CMC (NAEFS)**
  - **CPTEC**
  - **MeteoFrance**
  - **ECMWF**
  - **UKMO**
  - **WMO**

Definition of TIGGE database

- **The following details were agreed:**
  - **All accumulations to start from the beginning of the forecast**
  - **Geopotential Height to be used rather than Geopotential**
  - **Temperature extremes (max/min) to be provided over 6 hour intervals**
  - **Specific humidity to be provided in the free atmosphere**
  - **“2 metre temperature” to be used to refer to near surface temperature parameters**
  - **All fields to use units as defined in GRIB Edition 2**
  - **Orography and Land-sea mask to be provided for the Control for each output time-step.**
  - **Orography to be provided as geopotential height.**

List of products: single level

Parameter	Level	Unit	Output frequency	Comment
Mean sea level pressure	MSL	Pa	6h	instantaneous
Surface Pressure	surface	Pa	6h	inst
10m U-velocity	10m	m s**-1	6h	inst
10m V-velocity	10m	m s**-1	6h	inst
2m temperature	2m	K	6h	inst
2m dew point temperature	2m	K	6h	inst
2m max temperature	2m	K	6h	6_h
2m min temperature	2m	K	6h	6_h
Total precipitation (liquid + frozen)	surface	m	6h	acc_st
Snow fall	surface	m of water equivalent	6h	acc_st
Snow depth	surface	m of water equivalent	6h	inst

- **6_h: extrems over previous 6 hours**
- **acc_st: accumulated from start of forecast**

List of products: single level fields

Parameter	Level	Unit	Output frequency	Comment
Total cloud cover	surface	0-100%	6h	instantaneous
Total column water	surface	kg m**-2	6h	inst
Surface latent heat flux	surface	W m**-2 s	6h	acc_st
Surface sensible heat flux	surface	W m**-2 s	6h	acc_st
Surface solar radiation	surface	W m**-2 s	6h	acc_st
Surface thermal radiation	surface	W m**-2 s	6h	acc_st
Sunshine duration	surface	s	6h	acc_st
Convective available potential energy	surface	J kg**-1	6h	inst
Orography (Geopotential height at the surface)	surface	m		inst
Land-sea mask	surface	0-1		inst

- **acc_st: accumulated from start of forecast**
- **Orography and Land-sea mask to be provided for the Control for each output step**

List of products: upper air fields

Parameter	Unit	Output frequency	Comments
Temperature	K	6h	instantaneous
Geopotential height	m	6h	inst
U-velocity	m s**-1	6h	inst
V-velocity	m s**-1	6h	inst
Specific Humidity	kg kg**-1	6h	inst

- **5 parameters on 9 pressure levels, i.e. 45 fields.**
- **The 9 levels are 1000, 925, 850, 700, 600, 500, 300, 250 and 200 hPa.**

Data Policy

- **Data Providers to supply their products to the Archive Centres under an agreed set of rules, which will include re-distribution rights**
- **Access to be provided for Research & Education through a simple electronic registration process, with valid e-mail address and acknowledgment of conditions of supply**
- **Under the simple registration process, access to be given with a delay (e.g. 48 hours, to be defined) after initial time of the forecast**
- **Registration for real-time access to be handled via the THORPEX IPO**

Implementation plan

- **Test transfer rates between partners**
- **NCAR to investigate candidates for data transport**
- **GRIB2: ECMWF to consult with NAEFS and WMO Expert Team on Data Representation and Codes**
  - **Make sure that there is agreement on the proper encoding of the fields in GRIB2**
  - **Publish guidelines**
  - **Provide sample model output to the Data Providers**
- **Establish archive management communications:**
  - **Mailing lists, web sites and collaborative tools**
  - **Collect list of contact points**
- **Start feeding the TIGGE database …**

Conclusion

- **Milestones**
  - **2005-06: Initial infrastructure development**
  - **Early 2006: TIGGE data archives will begin collecting available global ensemble contributions in near-real time**
  - **2007-08: TIGGE available for THORPEX support to demo projects (IPY, Beijing 2008 Olympics regional EPS, …)**
- **Success of TIGGE depends on:**
  - **The commitment of each partners**
  - **The establishment of a collaboration methods**
  - **The availability of sufficient network bandwidth**
  - **The homogeneity of the catalogue that is built from standard metadata**