What are the user requirements in TIGGE?

The main task of the working group on infrastructure (WG-4) is to discuss the strategy for the technical implementation of TIGGE. In this context, technical implementation encompasses decisions on the data itself (volume, format, etc) as well as decisions on the technical infrastructure, which are necessary to facilitate TIGGE user requirements. Since these two aspects are inherently intertwined, the discussion necessarily will be an iterative process. In order to determine necessary resources and recommend the most appropriate infrastructure, WG-4 needs some input from the TIGGE users what are the actual requirements in terms of data volume, access etc. On the other hand, in order to avoid unrealistic demands, the TIGGE users need some guidance from the technical side on upper feasibility limits.

Below are some general questions and facts to be considered in the development of a common priority list of TIGGE user requirements.

- **What data should be exchanged and archived?**
  - how many parameters?
  - how many levels?
  - how many leveltypes?
  - what resolution?
  - how many ensemble members?
  - how many steps?
  - how many start dates per day?
  - how many models?

Some ballpark figures: A global field in T255 resolution is of the size of ~0.15 MB. Archiving ~50 ensemble members, at 60 forecast time-steps (00, 06, 12, 18 for 15 FC-days), 2 start dates per day, times ~60 fields (18 surface parameters, 7 parameters on 5 pressure levels, 7 parameters on 1 pv-level) would result in a data volume of ~54 GB.

0.15 MB x 50 = 7.5 MB (members)
7.5 MB x 60 = 450 MB (time-steps)
0.45 GB x 2 = 0.9 GB (start-dates)
0.9 GB x 60 = 54 GB (fields)

Note that in this exemplary calculation no model level data are included (which would be necessary for running LAMS). Adding them to the list of requirements would put a considerable strain on the systems.

Increasing the resolution to T399 would imply a factor of 2.5 for field sizes, resulting in a data volume of 135 GB per day! Assuming 10 EPS producing centres in TIGGE, but reduced resolution and less ensemble members for them, additional data of 3 times the ECMWF volume might be realistic. This would result in a data volume of ~400 GB per day, or ~150 TB per year.

To receive 400 GB per day via the internet, 37 Mbps sustained throughout the day are necessary. To allow for variability, this will require a 100 Mbps link (currently, ECMWF has 60 Mbps internet link). If the above calculated data volume is really necessary (or even more data e.g. model levels have to be added), shipping tapes once per month might be the only way to achieve this. The research community has to decide whether data volume or immediate availability is of priority.

The projected ECMWF archive growth for 2005 is ~1.4 PB; TIGGE data would be ~10% of the primary data growth. Thus, the main infrastructure challenge seems to be the data exchange and not so much the data archive.

There are obviously ways to reduce or increase the above given data volume estimate by reducing/increasing the components of the calculation (number of start dates, time-steps, fields, resolution to be archived). Thus, the TIGGE users have to agree to a priority-list for the TIGGE archive.

- **What data should be available for "quick access" / frequent exchange**
  Taking into account the above-mentioned constraints, one option could be to define a data subset available for "quick access".

- **What time-delay for data provision ("quick-access" / normal access) is acceptable?**
  It has to be decided how "operational" the support for the "quick-access" data has to be, and how long the research community can wait for e.g. "shipped" data.
- **What will be the typical request: complete fields, sub-areas, time series?**
  In order to design the most efficient data archive structure a list of typical requests, ranging from the most popular to very infrequent requests, has to be provided to WG-4.
- **What data formats will have to be supported (GRIB(2), BUFR, NetCDF,…)**
  The more data formats have to be supported, the more expensive it gets (person-power, hardware/software costs). On the other hand, offering the data in various formats would make the data more popular and enlarge the potential user community. A trade-off between the extremes has to be found.
- **What tools and applications are needed?**
  It has to be decided what tools and applications are essential and should be centrally provided. Examples are: common data interface, basic data manipulation tools (averaging etc), plotting packages…

Here we need your input!

In order to produce a first draft priority list of TIGGE user requirements we need your input! Please have a look at the following very specific question and send us your answers/comments as soon as possible. We'll compile the response and present the resulting draft priority list at the workshop for further discussion. This should also give WG-4 a good starting point for discussing the design of the infrastructure.

Data exchange:

1. Do you think that the data exchange strategy should be the same for the whole content of the archive, or do you prefer the concept of defining a subset of the archive for quick access/exchange on a daily basis?
2. What data would you need in the "quick exchange" data subset?
3. What is the maximum acceptable time-delay for accessing "quick" or "normal" data?

Data archive:

4. Please give a ranking of the potential content of the TIGGE archive and sort the following factors according to priority, i.e. is it most important to you that:
   - all (or x) ensemble members are archived?
   - data are archived on the original resolution?
   - X surface parameters are archived?
   - X parameters on Y pressure levels are archived?
   - X time-steps per day are archived?
   - X start dates per day are archived?
   - X parameters on Y model levels are archived?
5. What will be your typical request to the archive? Please give an indication where it lies between the two extremes of global fields or local time-series?
6. What is your preferred data format? Do you need more than one format (please give list of priorities)?

Tools and applications:

7. What tools do you need or find useful? Please give a priority list of tools and/or applications which should be made available.
8. Would you find the possibility of your own application to be installed at the archive centre useful?